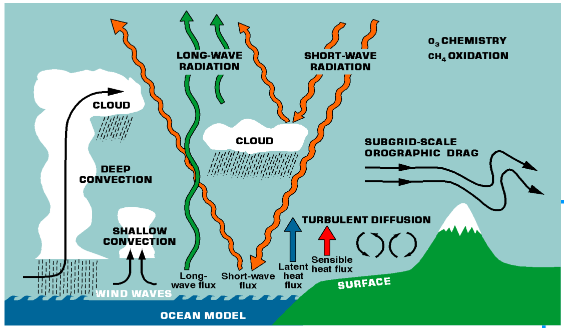
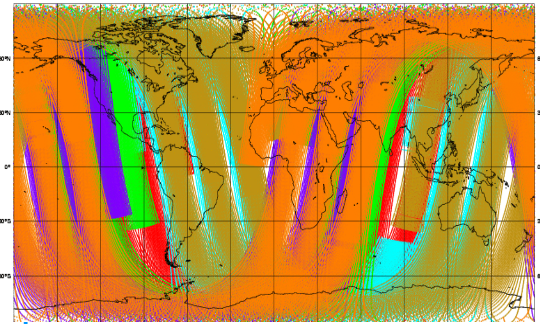
Data assimilation at ECMWF

Massimo Bonavita
ECMWF
Data Assimilation Section
massimo.bonavita@ecmwf.int

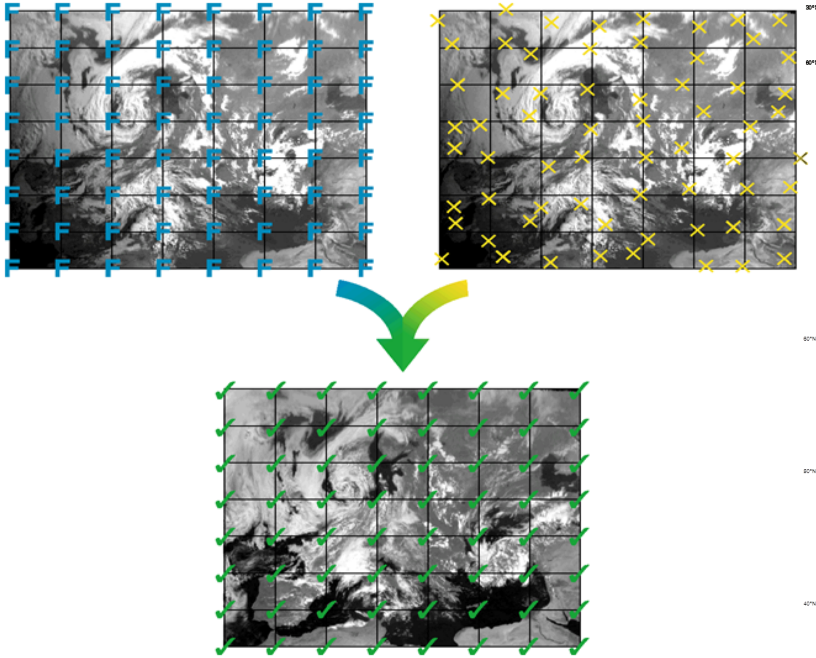
Data assimilation at ECMWF

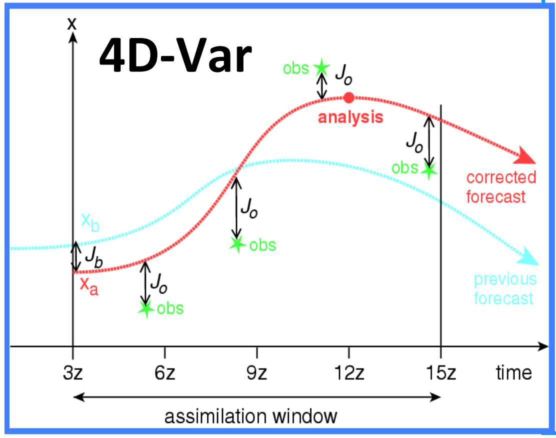


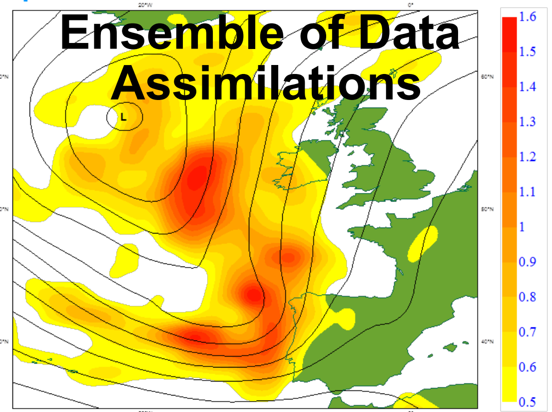
Forecast model

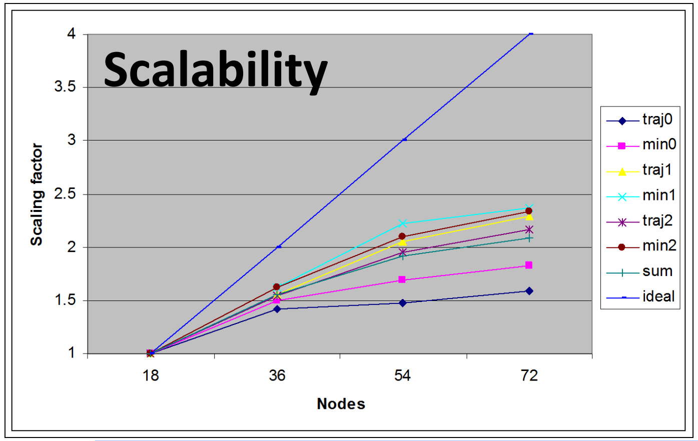


Observations



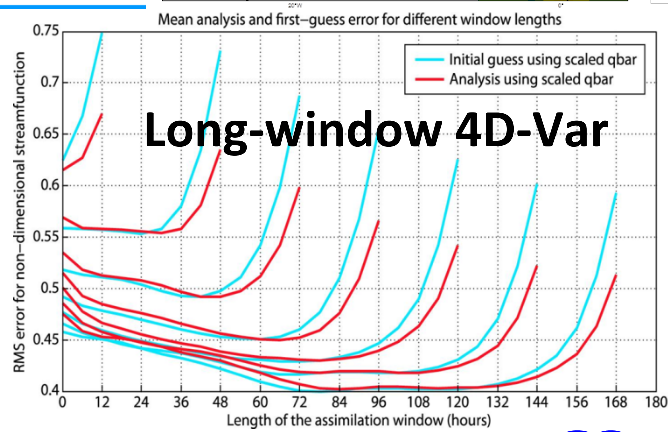






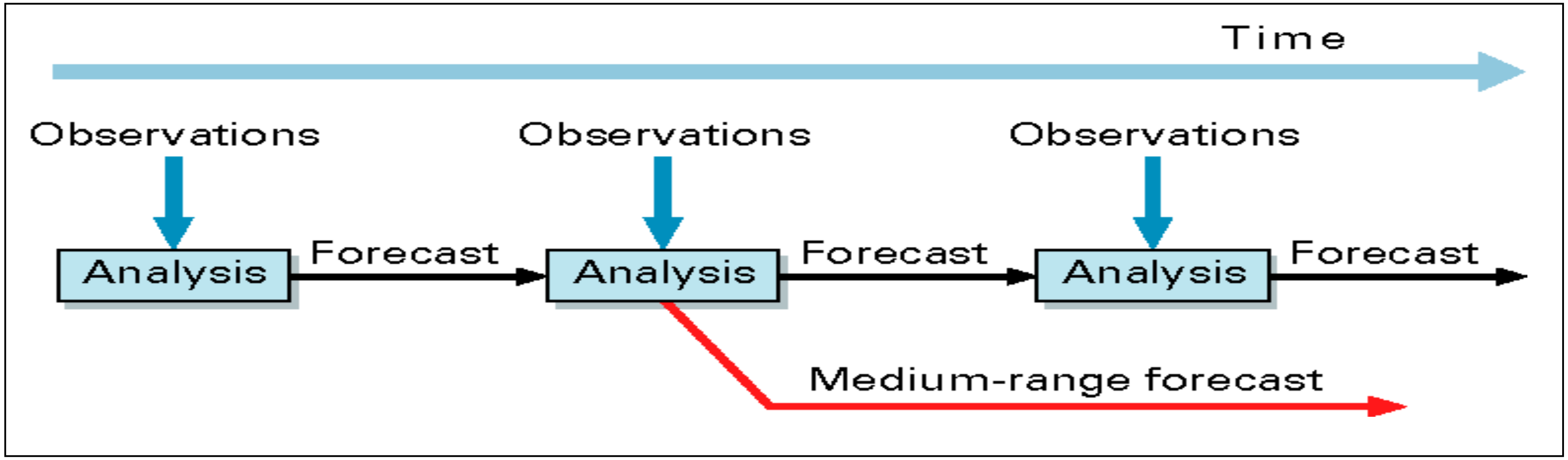
Methods

Progress and plans



ECMWF

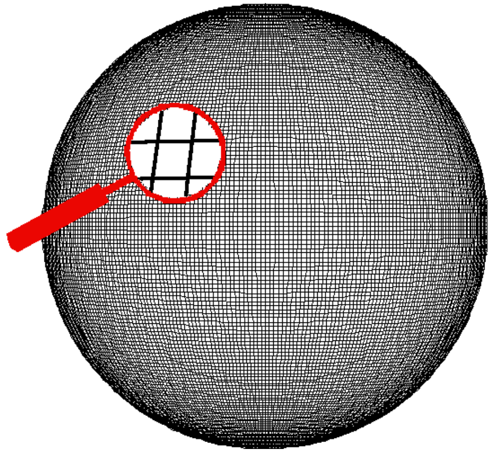
Data assimilation

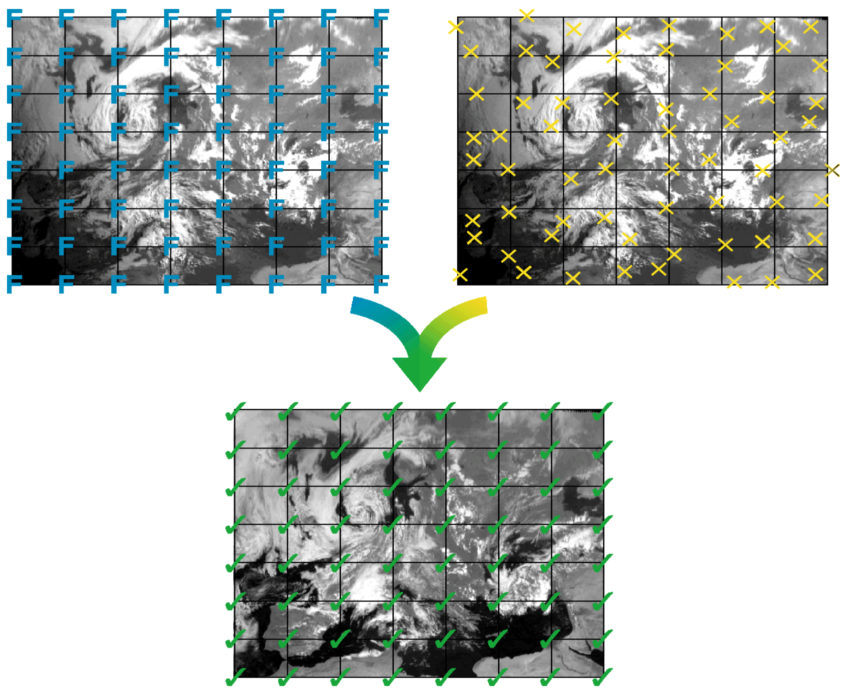


- The observations are used to correct errors in the short forecast from the previous analysis time.
- At ECMWF, twice a day 15 – 16,000,000 observations are used to correct the 80,000,000 variables that define the model's virtual atmosphere.
- This is done by a careful 4-dimensional interpolation in space and time of the available observations; this operation takes as much computer power as the 10-day forecast.

Data assimilation for weather prediction

The **FORECAST** is computed on a quasi-regular grid over the globe.
The meteorological **OBSERVATIONS** come from any location on the globe.
The computer model's prediction of the atmosphere is compared against the available observations, in near real time





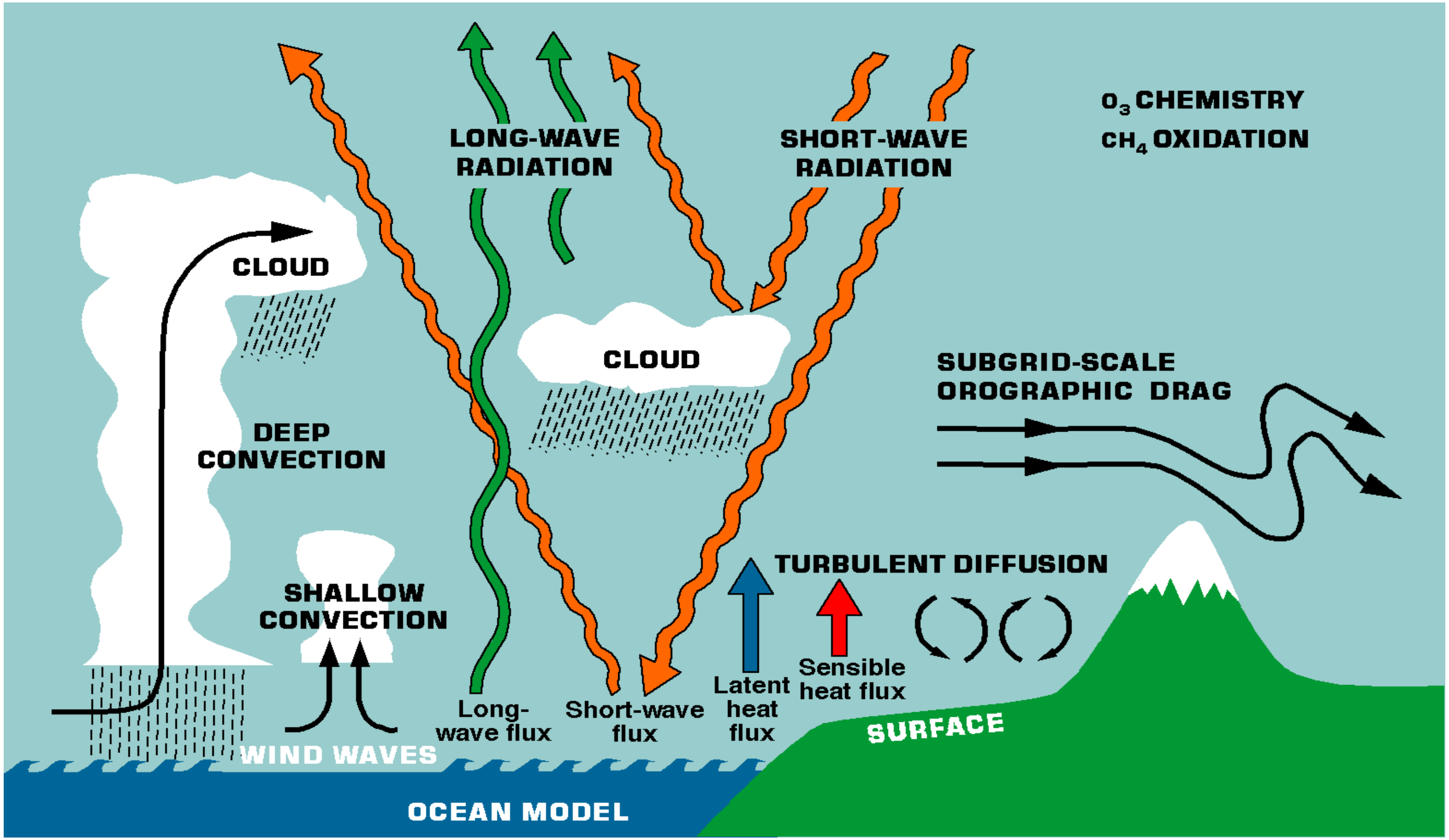
A short-range **forecast** provides an estimate of the atmosphere that is compared with the **observations**.

The two kinds of information are combined to form a corrected atmospheric state: the **analysis**.

Corrections are computed and applied twice per day. This process is called '**Data Assimilation**'.

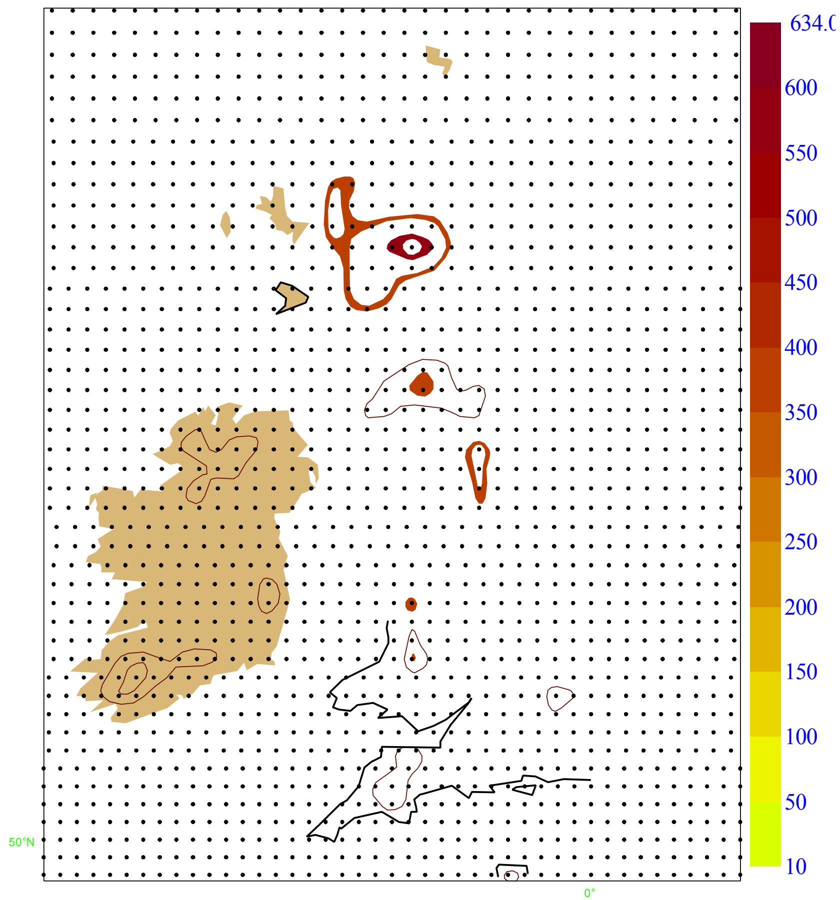
The ECMWF forecast model is a very important component of the data assimilation system

Physical processes in the ECMWF model



T_L799

Previous operational resolution

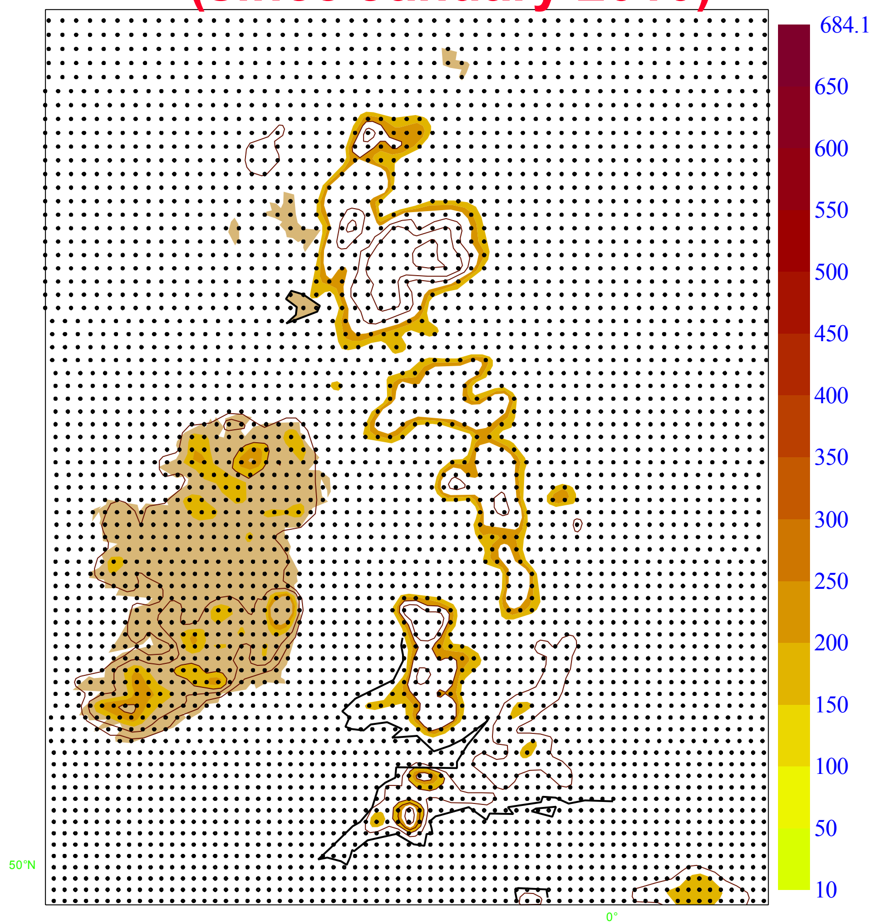


25 km grid-spacing

(843,490 grid-points)

T_L1279

Current operational resolution (since January 2010)

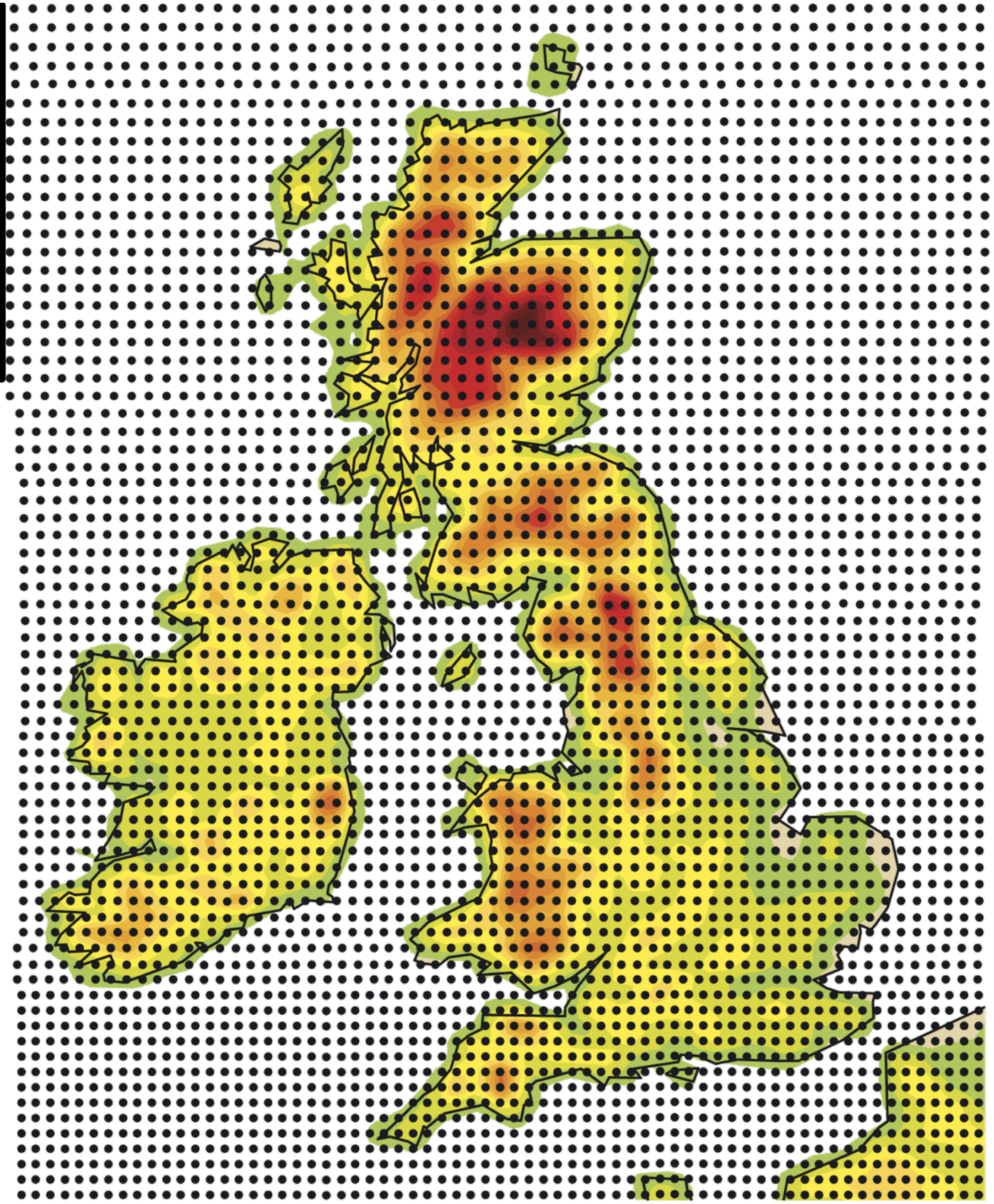


16 km grid-spacing

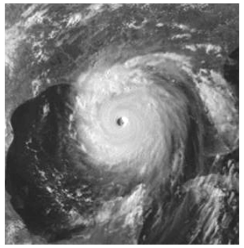
(2,140,704 grid-points)

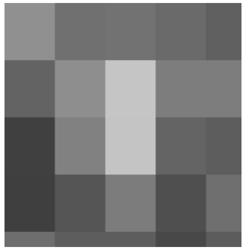
Model resolution matters.

Present ECMWF system:
Global model with 16 km resolution and 137 levels

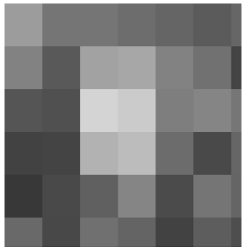


Increasing Resolution

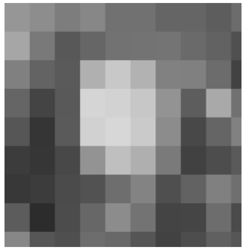




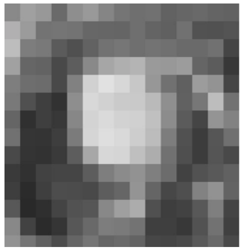
~210km



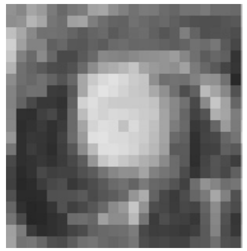
~125km



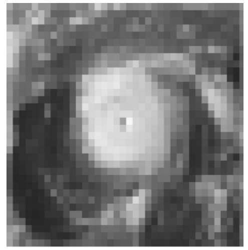
~63km



~39km

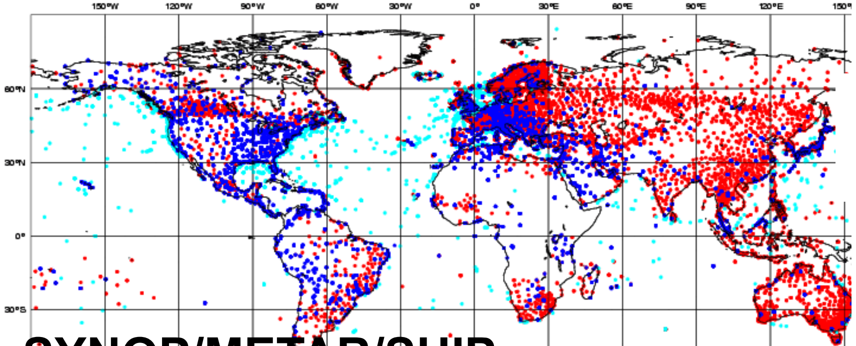


~25km

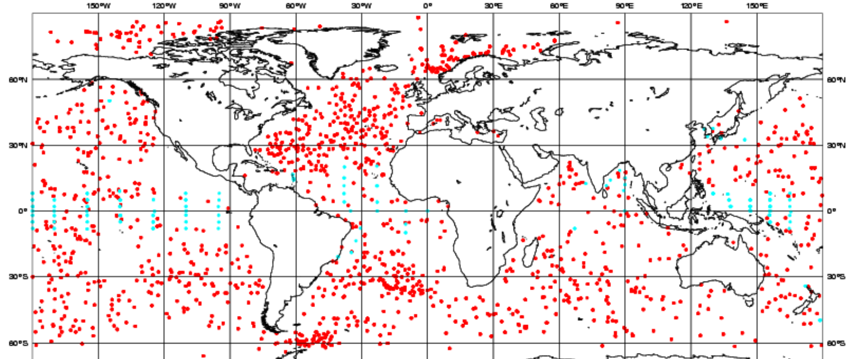


~16km

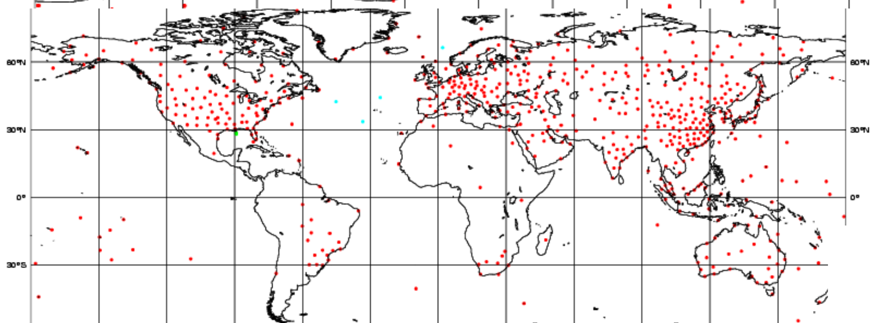
Conventional observations used by ECMWF's analysis



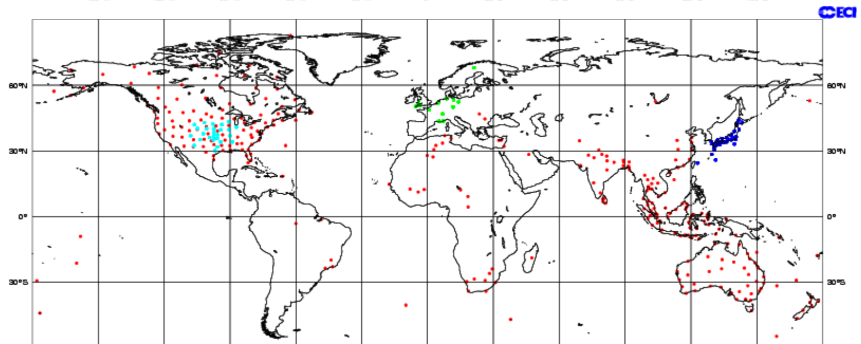
SYNOP/METAR/SHIP: MSL Pressure, 10m-wind, 2m-Rel.Hum.



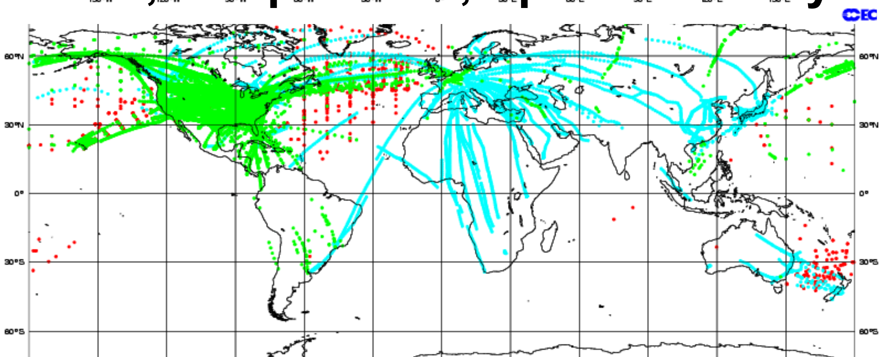
DRIBU: MSL Pressure, Wind-10m



Radiosonde balloons (TEMP): Wind, Temperature, Spec. Humidity



PILOT/Profilers: Wind

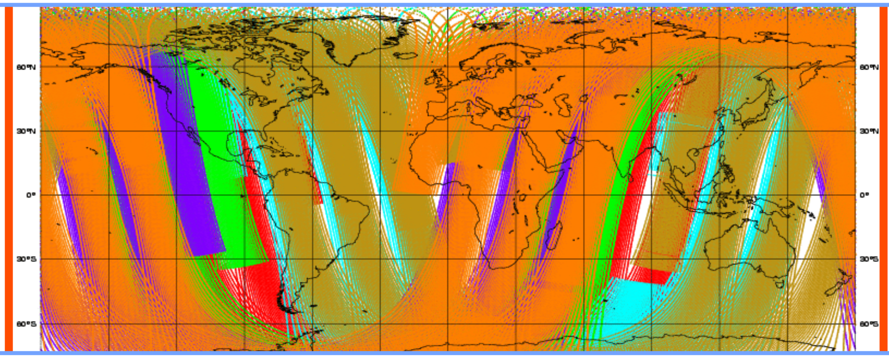


Aircraft: Wind, Temperature

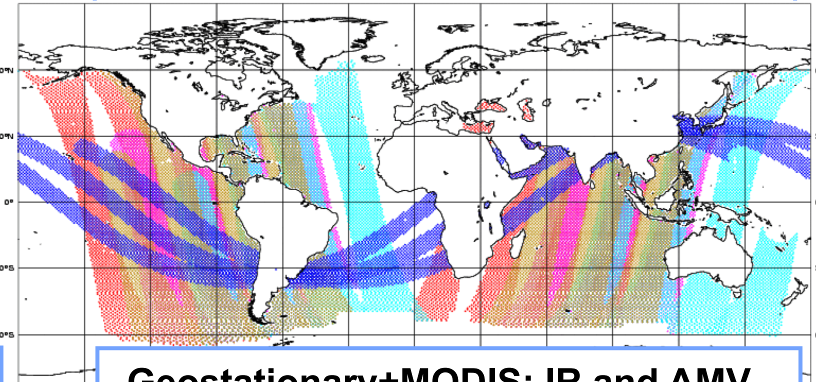
Note: We only use a limited number of the observed variables; especially over land.

Satellite data sources used by ECMWF's analysis

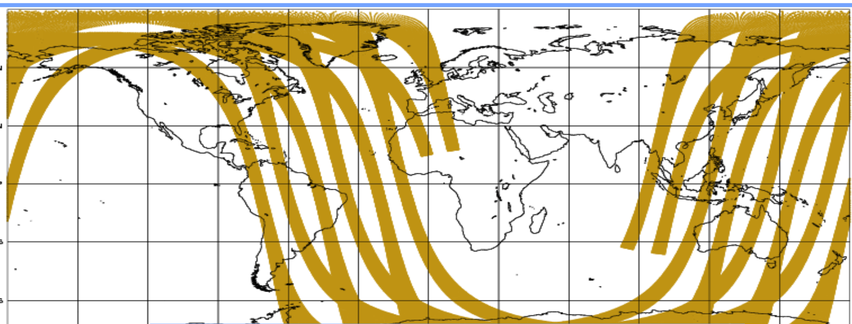
Sounders: NOAA AMSU-A/B, HIRS, AIRS, IASI, MHS



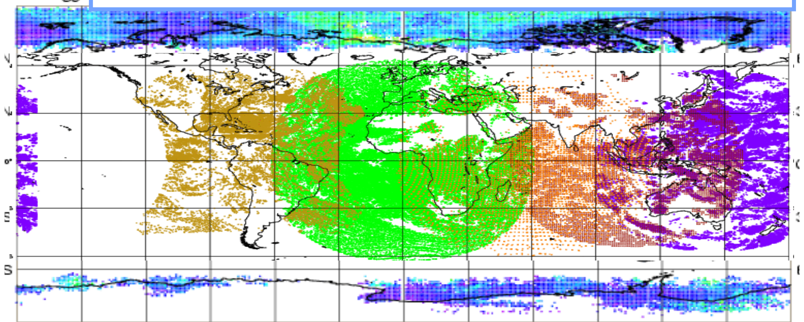
Imagers: SSMI, SSMIS, AMSR-E, TMI



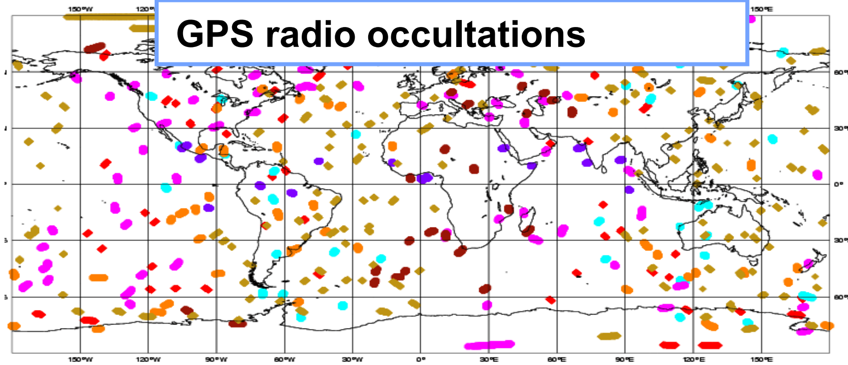
Scatterometer ocean low-level winds: ASCAT



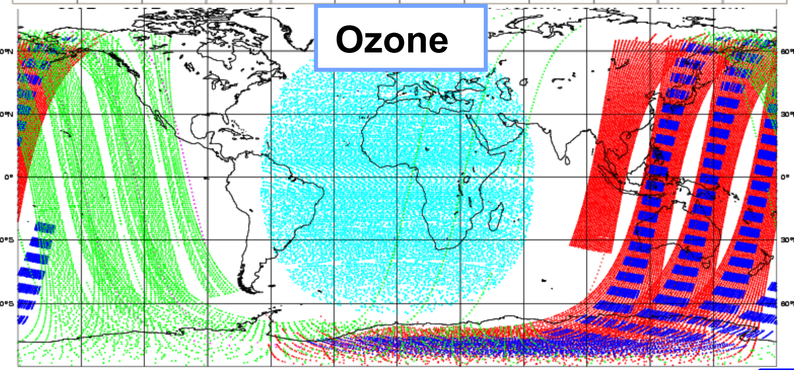
Geostationary+MODIS: IR and AMV



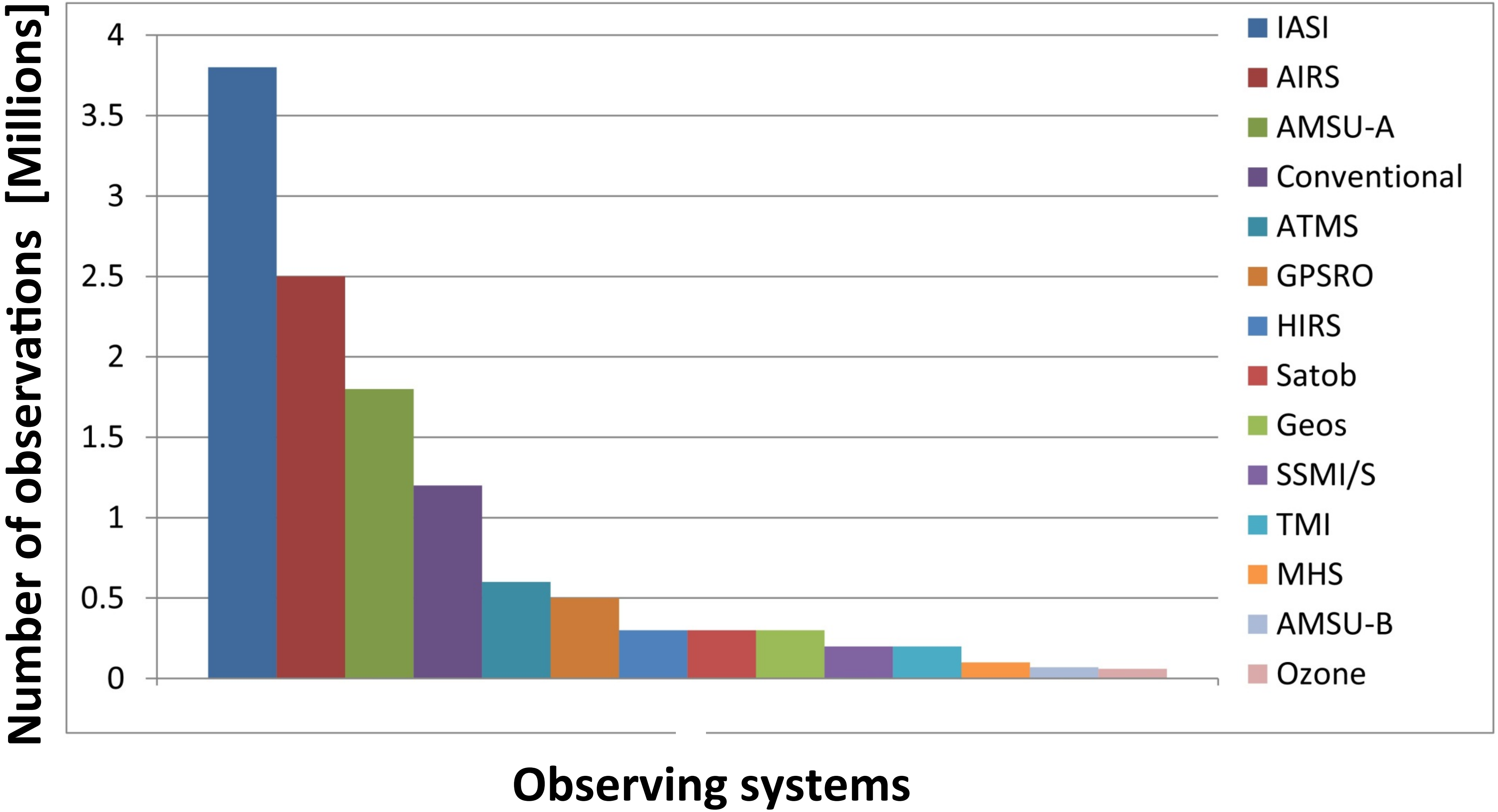
GPS radio occultations



Ozone

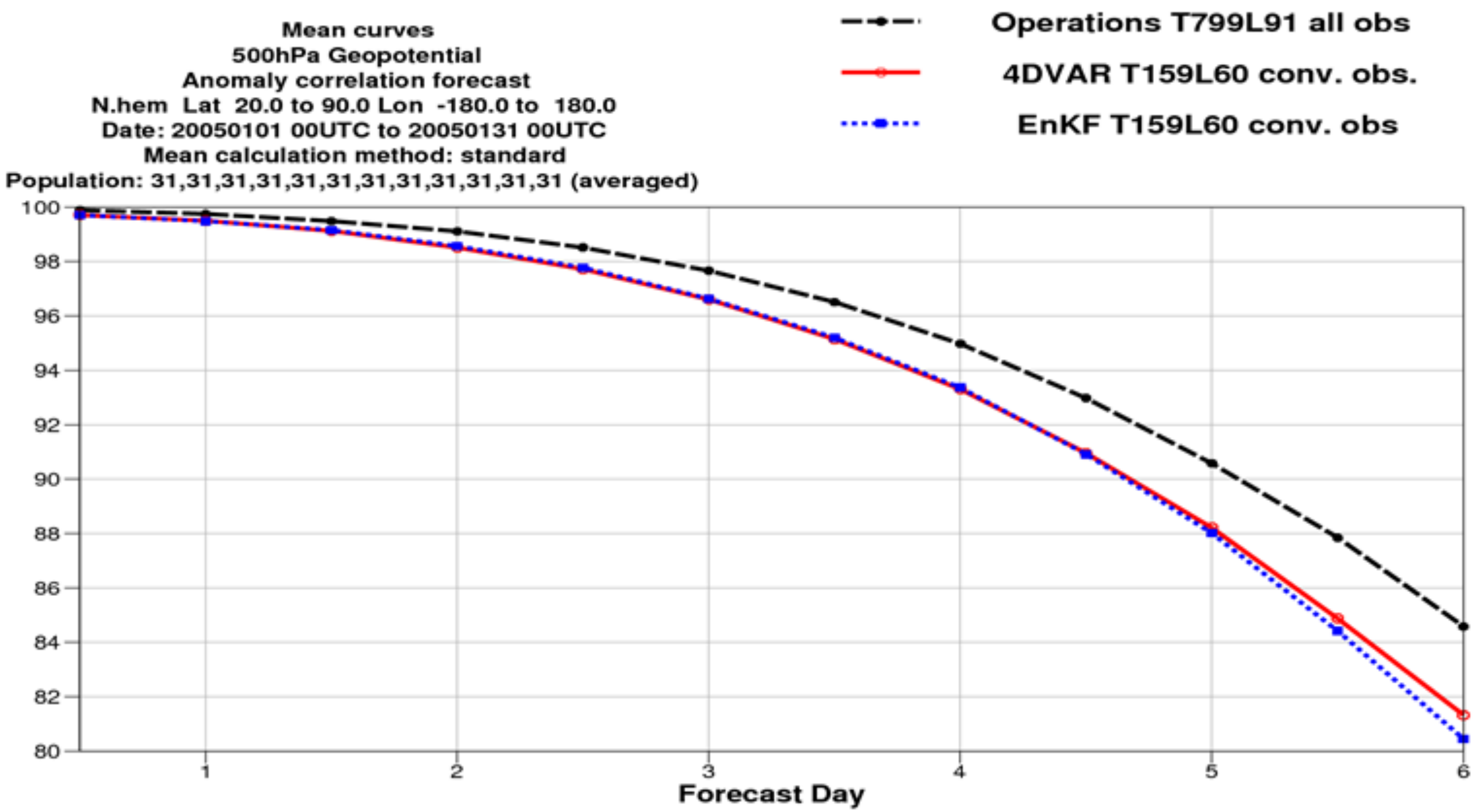


Number of observations used for a 12-hour 4D-Var analysis: Total approx. 15M (conventional data approx. 1.2M)



Observations

- Satellite Obs. make up ≈ 95% of total used observations
- Conventional Obs. are still very important in the North Hem. and for the bias correction of the satellite radiances



Quality control of observations is very important

Data extraction

- Check out duplicate reports
- Ship tracks check
- Hydrostatic check

Thinning

- Some data is not used to avoid over-sampling and correlated errors
- Departures and flags are still calculated for further assessment

Blacklisting

- Data skipped due to systematic bad performance or due to different considerations (e.g. data being assessed in passive mode)
- Departures and flags available for all data for further assessment

Model/4D-Var dependent QC

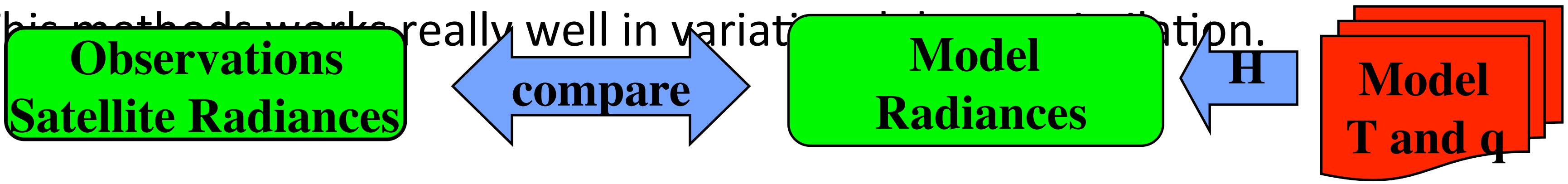
- First guess based rejections
- VarQC rejections

Used data → **Increments**

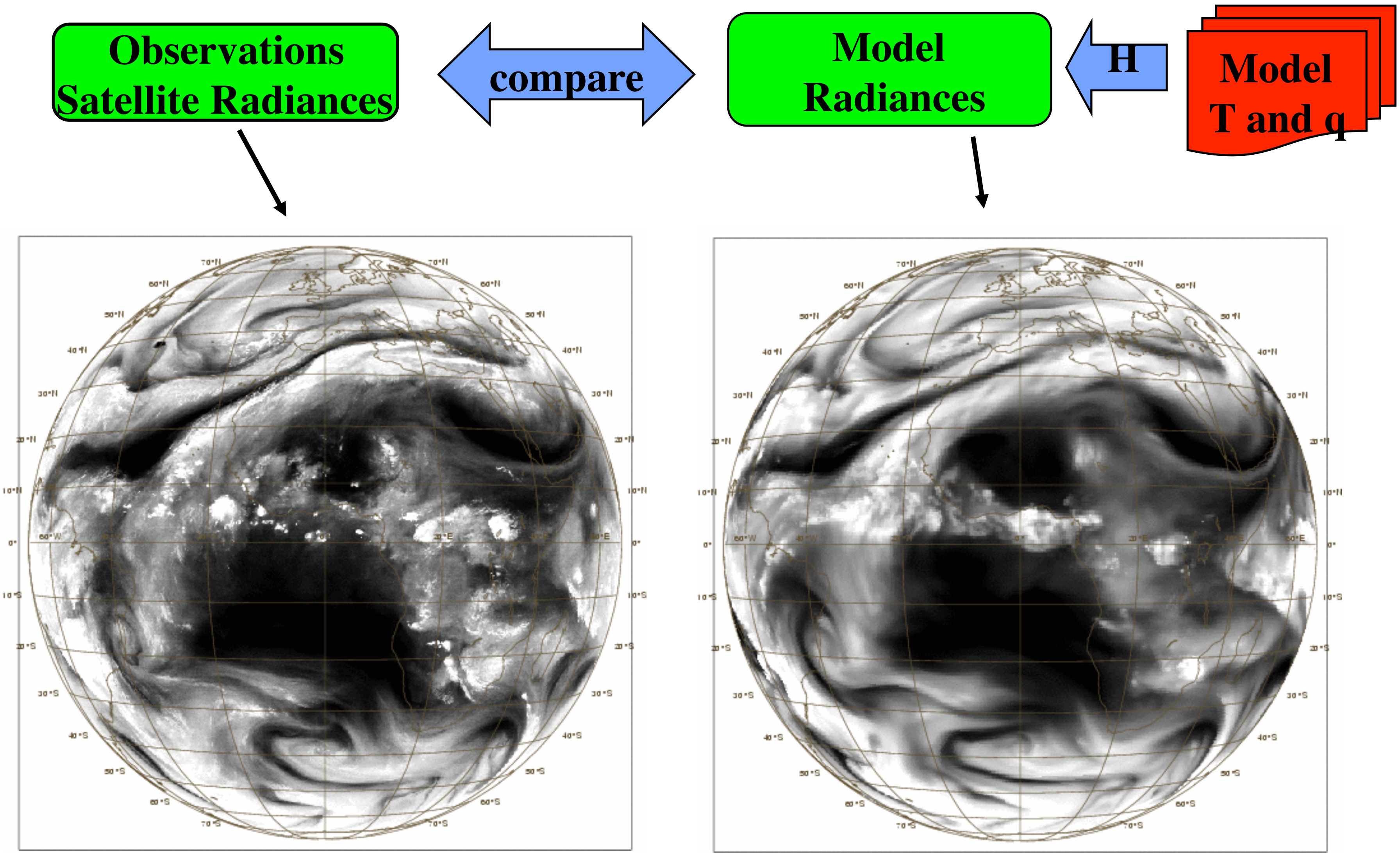
Analysis

How we use satellite data in the analysis

- Observations are not made at model grid points.
- Satellites measure radiances, NOT temperature and humidity .
- We calculate a model radiance estimate of the radiance measurement, using a so-called 'observation operator' **H**.
- **H** performs a complex transformations of model variables (T,q,O3) to radiances.
- The model estimate is compared with the observed radiance.
- The difference between the observed radiance and the radiance estimated by the model (background departure) is used in the analysis algorithm.
- This methods works really well in variational data assimilation.

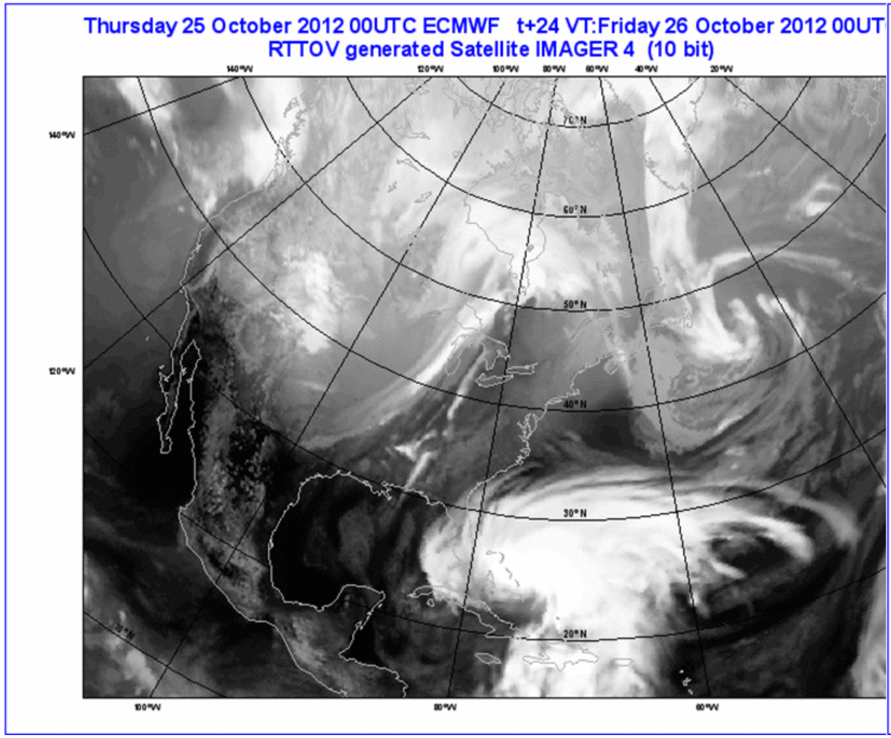


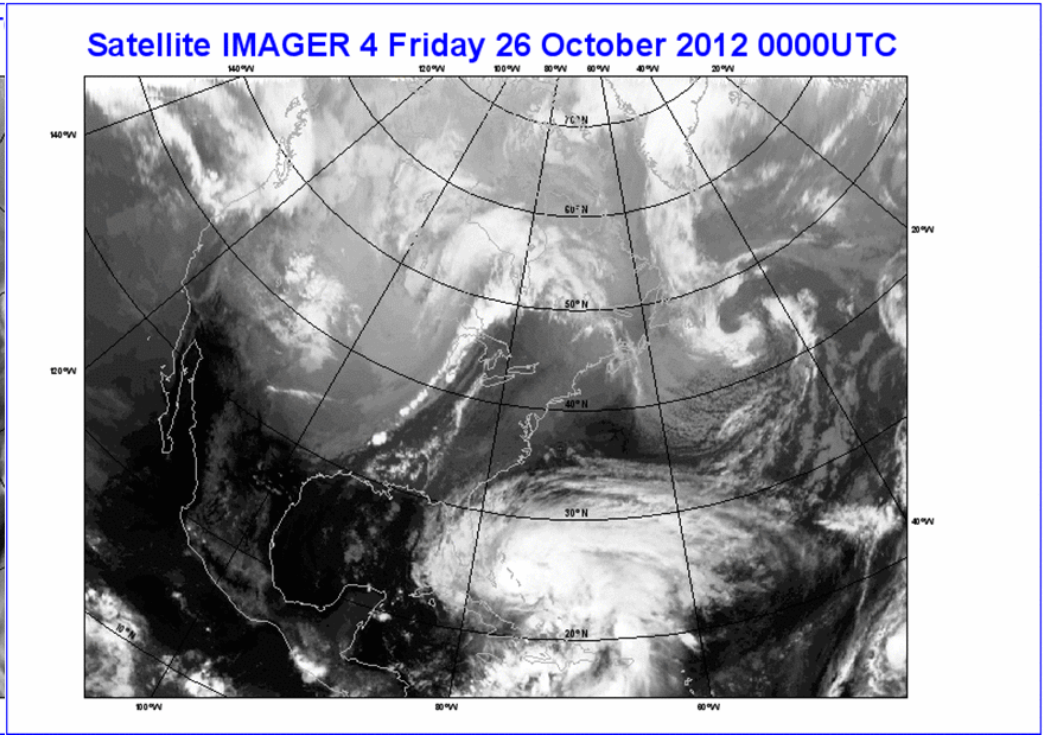
The variational method allows model radiances to be compared directly to observed radiances



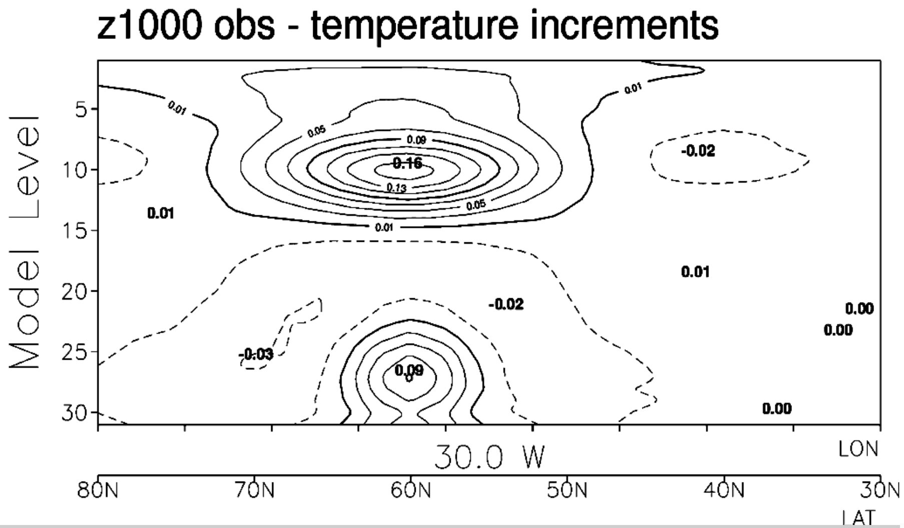
ECMWF

Hurricane Sandy 22-30 Oct. 2012

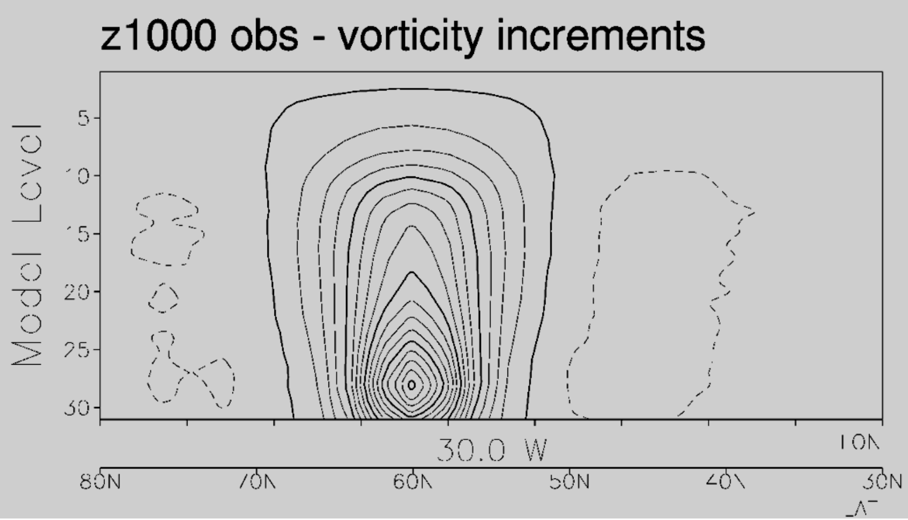


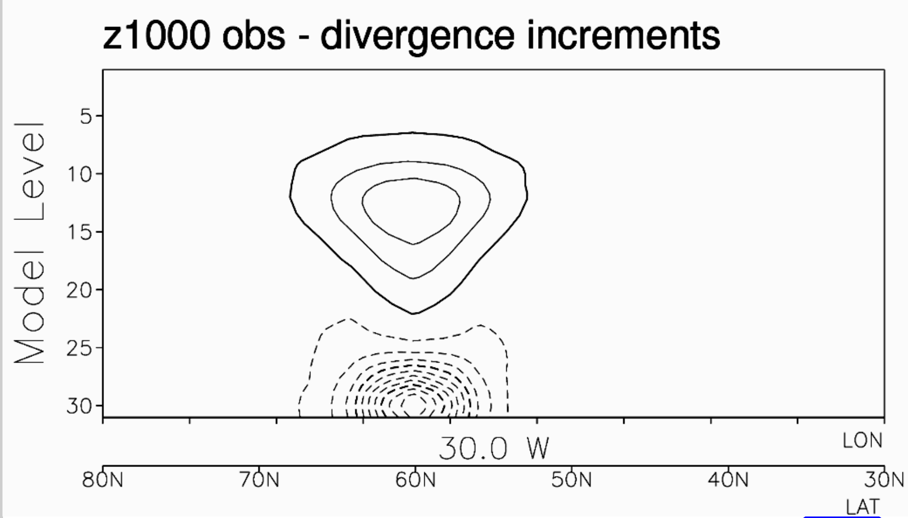


Analysis corrections are meteorologically consistent!



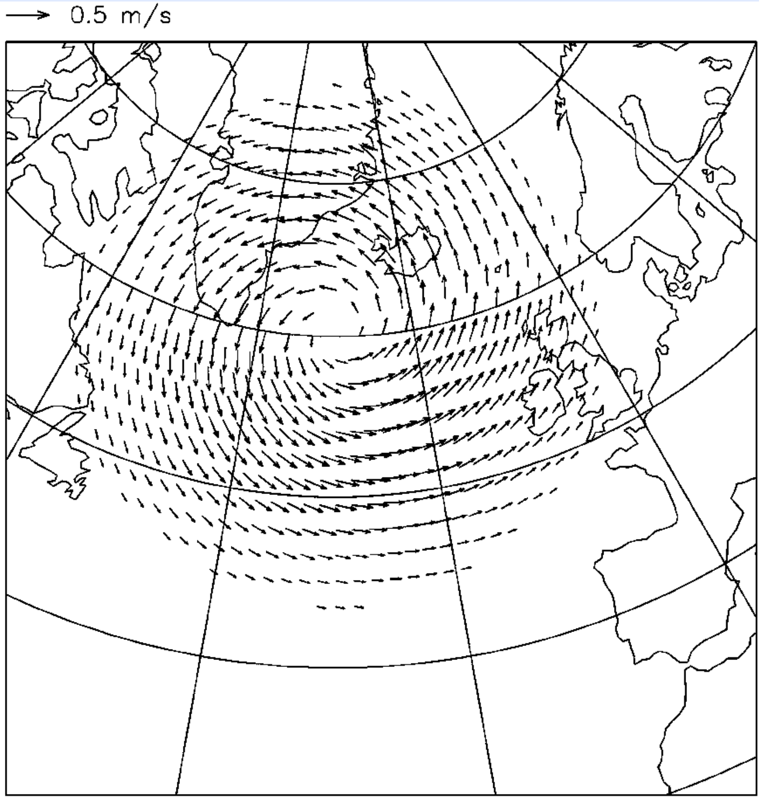
Increments due to a single observation of geopotential height at 1000hPa at 60N with value 10m below the background.



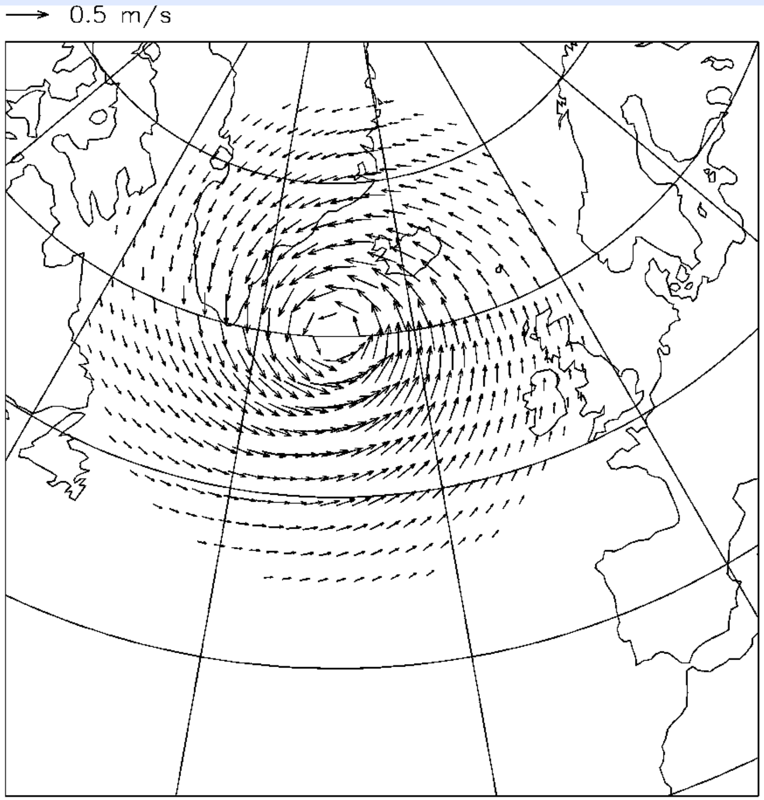


Analysis corrections are meteorologically consistent! Height and wind field balance is retained in the extra-tropics

wind increments at 300hPa



wind increments 150 metre above surface

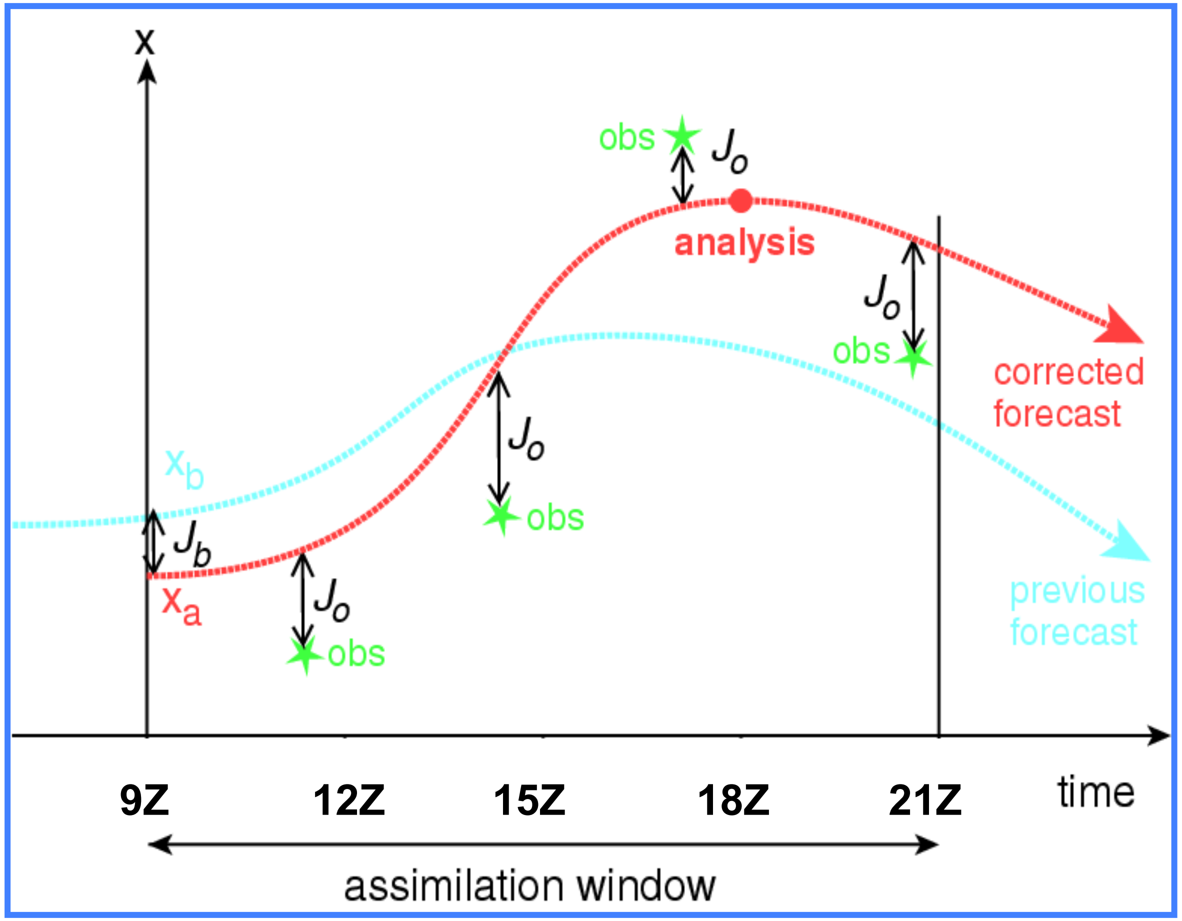


Increments for a single observation of geopotential height at 1000hPa.

A few 4D-Var Characteristics

- All observations within a 12-hour period (~15,000,000) are used simultaneously in one global (iterative) estimation problem

- "Observation – model" values are computed at the observation time at high resolution: **16 km**
- 4D-Var finds the 12-hour forecast that fits the observations in a dynamically consistent way.
- Based on a tangent linear and adjoint forecast models, used in the minimization process.
- 80,000,000 model variables (surface pressure, temperature, wind, specific humidity and ozone) are adjusted



Incremental 4DVar

The 4DVar analysis is found by an iterative process in which the difference between the short range forecast model and the observations is minimised in the 12h assimilation window

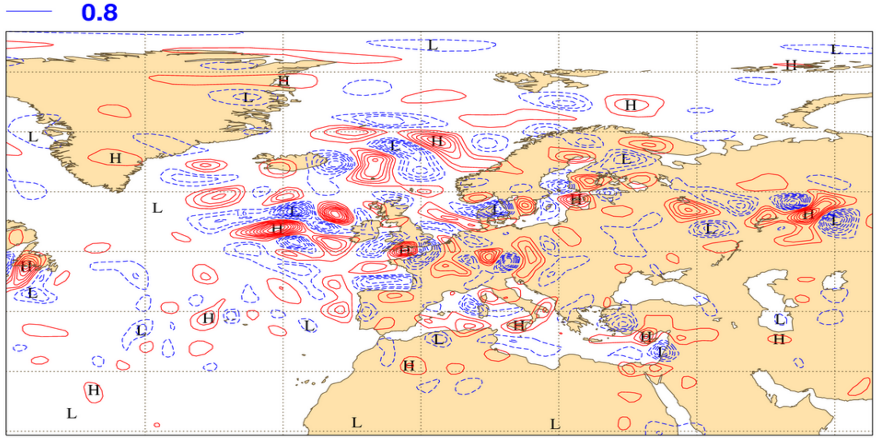
To reduce the computational cost of 4DVar the minimization is performed at lower resolution , in an incremental manner

Incremental 4DVar

Analysis increments for vorticity, 500 hPa, 2012/09/30 21UTC

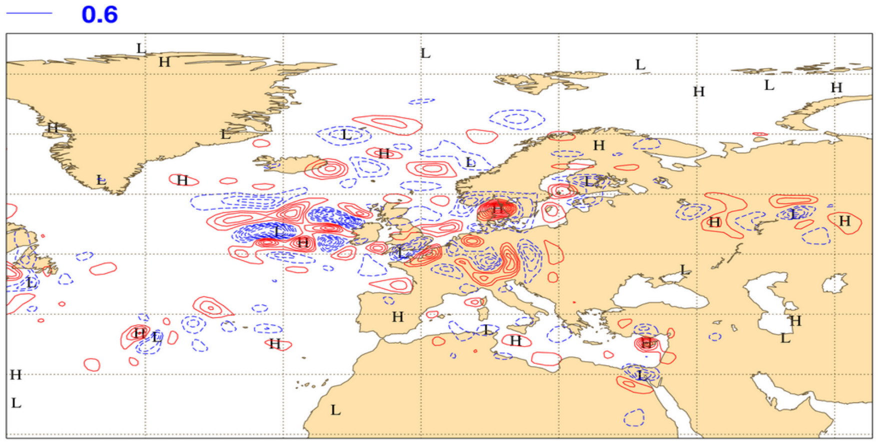
1st Minimization

T159 (~125 Km)



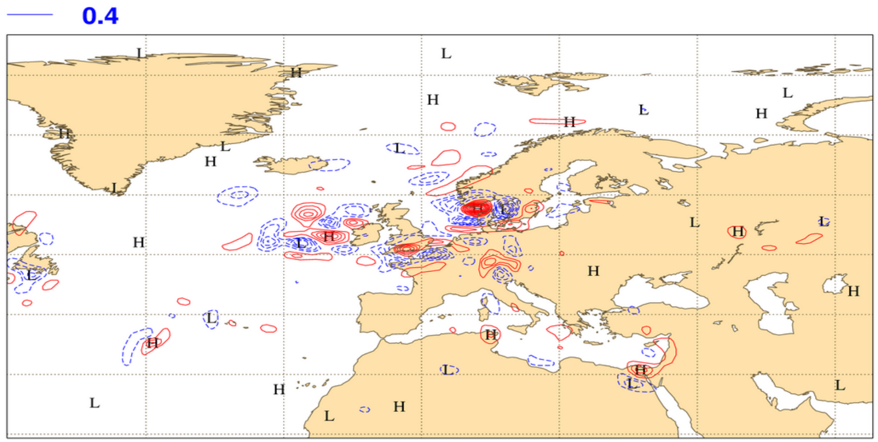
2nd Minimization

T255 (~80 Km)



3rd Minimization

T255 (~80 Km)

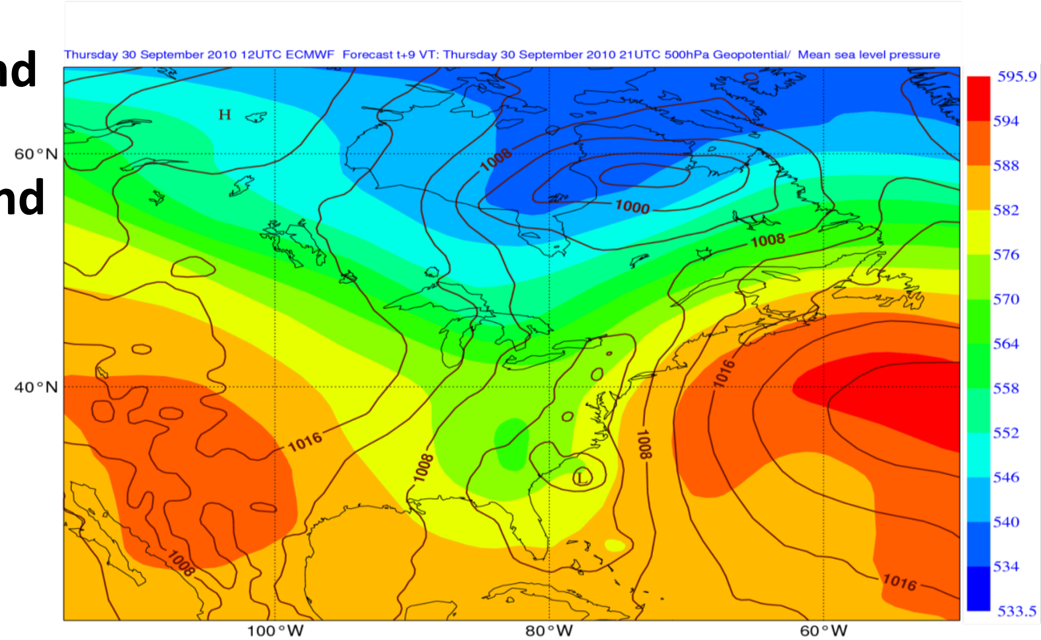


Incremental 4DVar

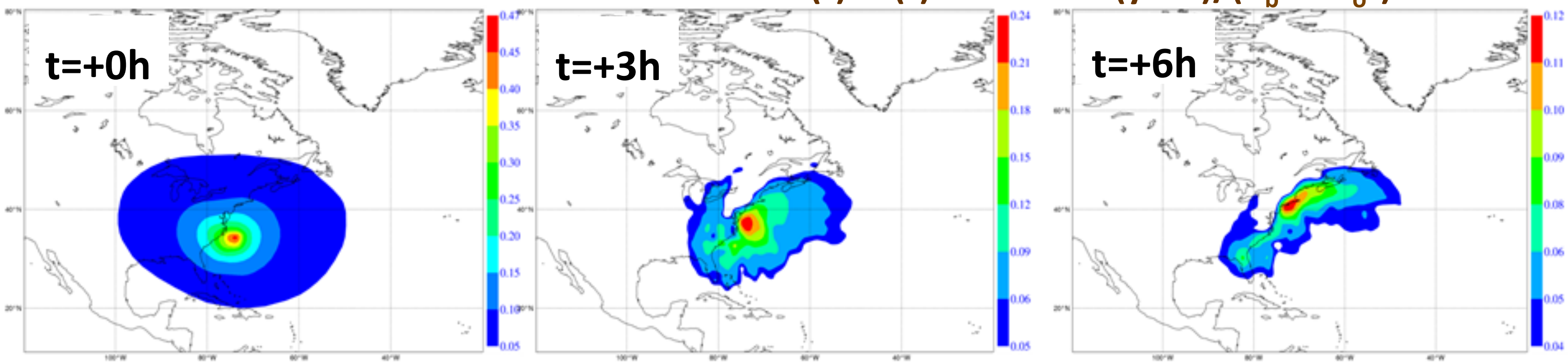
A useful property of 4DVar is that it implicitly evolves the analysis increments *over the length of the assimilation window* (*Thepaut et al*.,1996) in accordance with the model dynamics

Incremental 4DVar

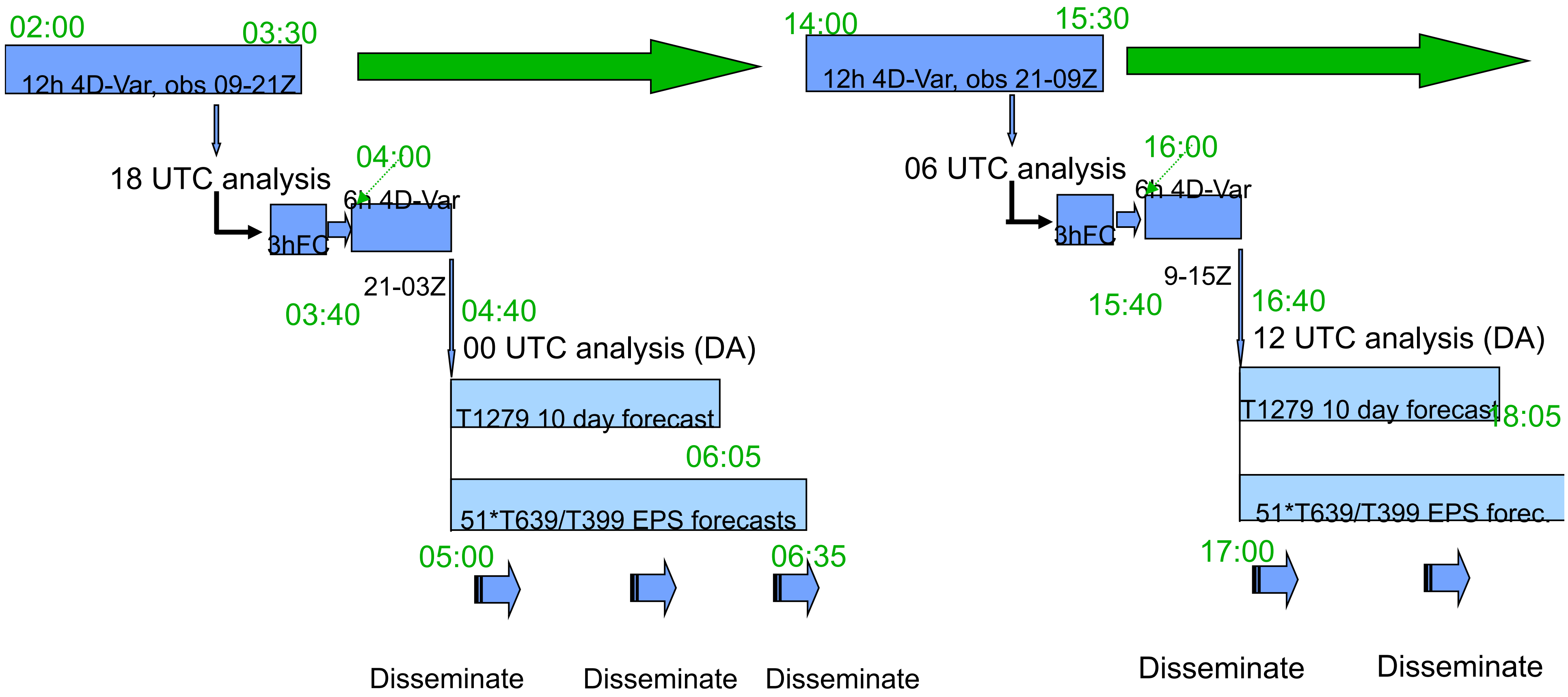
MSLP (contours) and 500 hPa height (shaded) background



Temperature analysis increments for a single temperature observation at the start of the assimilation window: $x^a(t)-x^b(t) \approx \mathbf{MBM^TH^T}(y-Hx)/(\sigma_b^2 + \sigma_o^2)$



Operational schedule
Early delivery suite introduced June 2004



ECMWF

Recent operational data assimilation changes at ECMWF

- Improved humidity analysis, accounting better for super saturation effects
- Improved scalability – but still more to be done
- Reduced observation error for AMSU-A radiances
- Bias correction of aircraft temperature observations
- Using the data assimilation system for ERA-20C (1900-2010)
- Assimilation of rain-affected microwave satellite data

- Use EDA to provide flow-dependent background error variances in 4D-Var
- Use EDA to provide flow-dependent background error covariances in 4D-Var
- Extended Kalman Filter (EKF) for soil moisture analysis
- New snow analysis and higher resolution snow satellite data

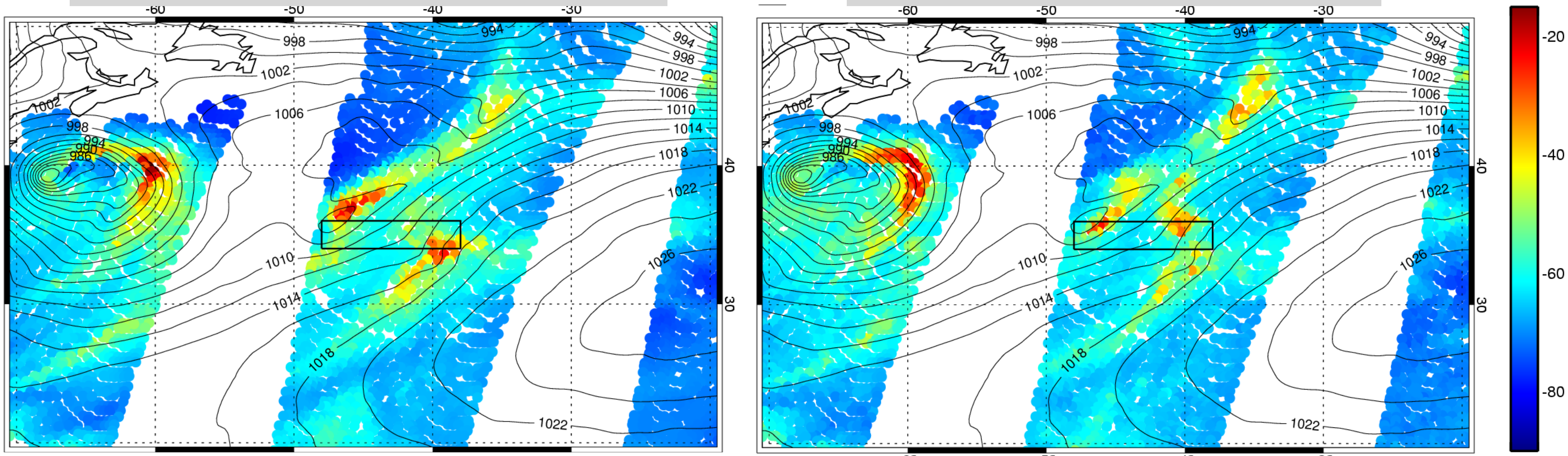
Improved assimilation of satellite moisture data

Assimilation of rain-affected microwave

- First version (SSM/I radiances) 2005; extended to SSMIS, TMI, AMSR-E in 2007
- Direct 4D-Var radiance assimilation from March 2009; improved 2010; improved 2011
- Main difficulties: inaccurate moist physics parameterizations (location/intensity), formulation of observation errors, bias correction, linearity.

4D-Var first guess SSM/I ΔTb 19v-19h [K]

SSM/I observational ΔTb 19v-19h [K]

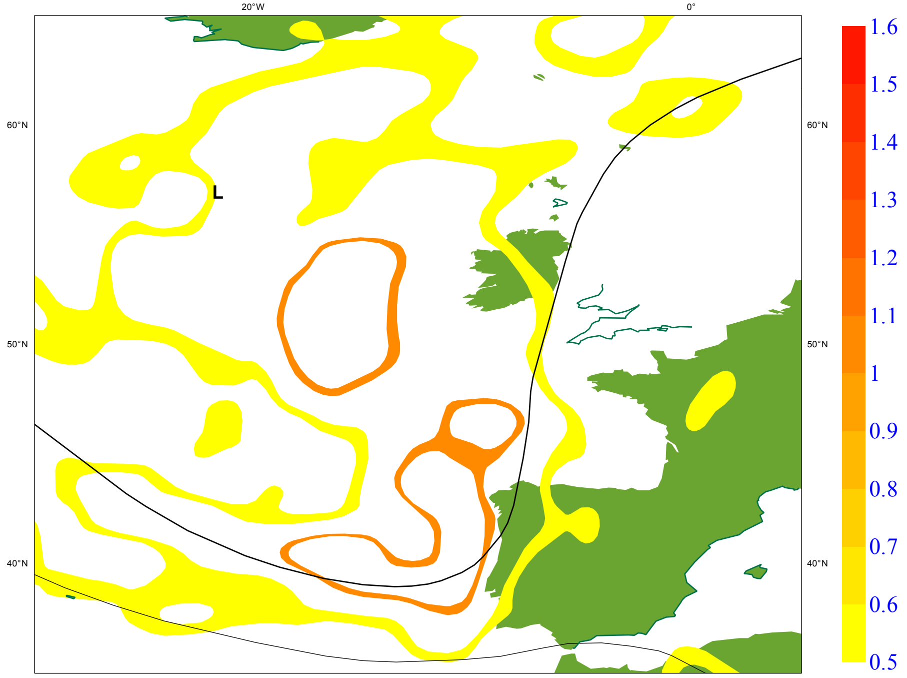


ECMWF

Ensemble of Data Assimilations (EDA)

- Run an ensemble of analyses with perturbed observations, perturbed model physics and perturbed Sea Surface Temperature fields.
- 25 EDA members plus a control at lower resolution.
- Form differences between pairs of analyses (and short-range forecast) fields.
- These differences will have the statistical characteristics of analysis (and short-range forecast) error.

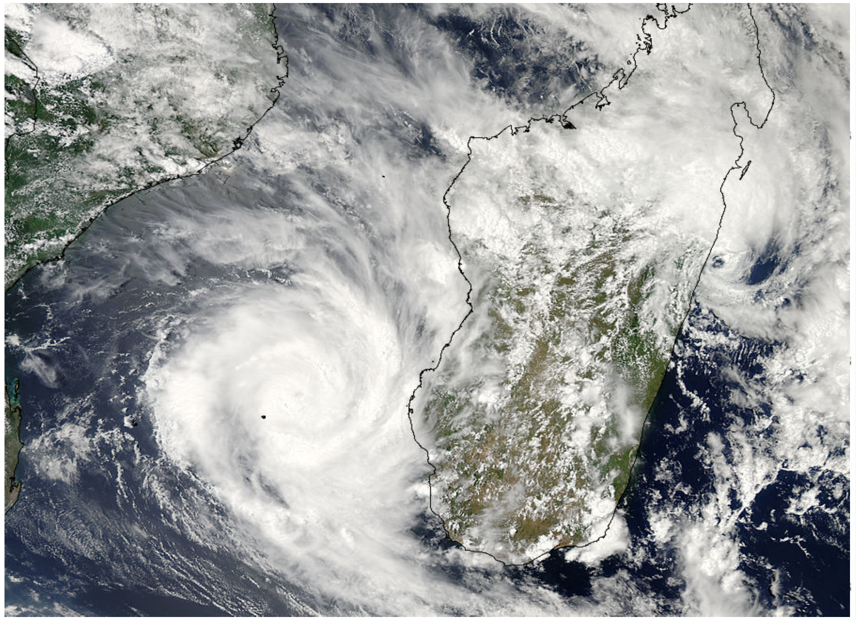
Yellow shading where the short-range forecast is uncertain: This will give observations more weight in these regions.



ECMWF

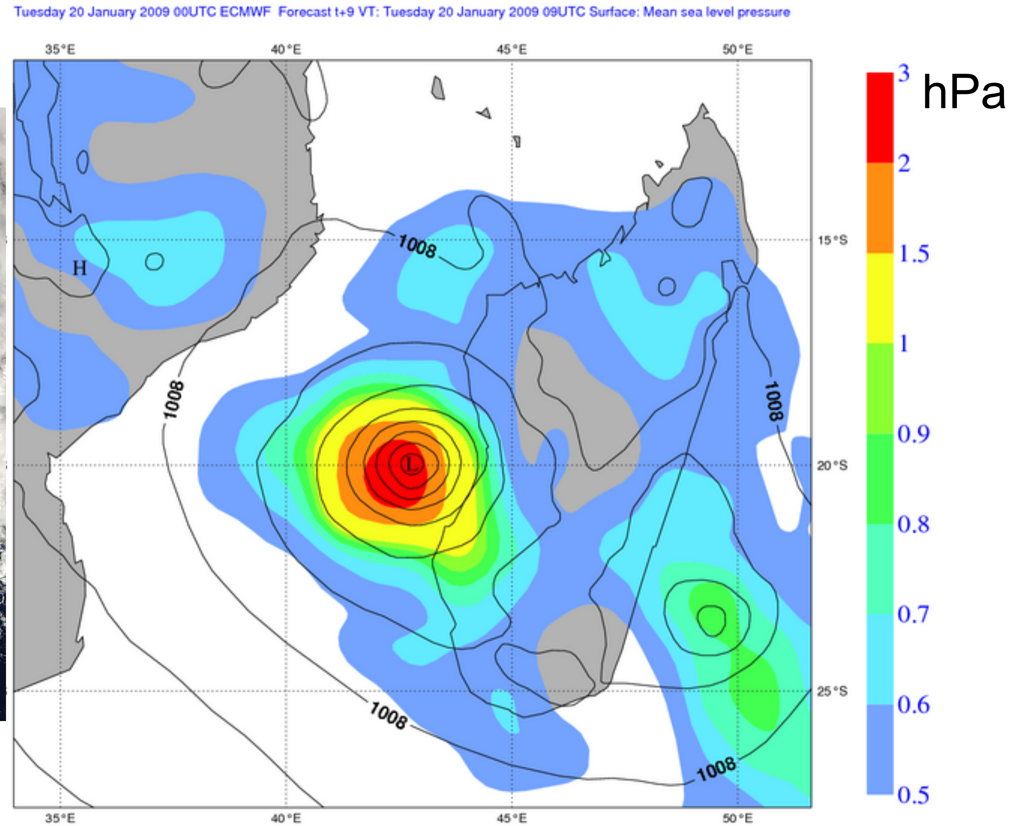
In May 2011 ECMWF implemented EDA based flow-dependent background error variance in 4D-Var

The 10-member EDA has been used to estimate the background error variance in the deterministic 4D-Var.



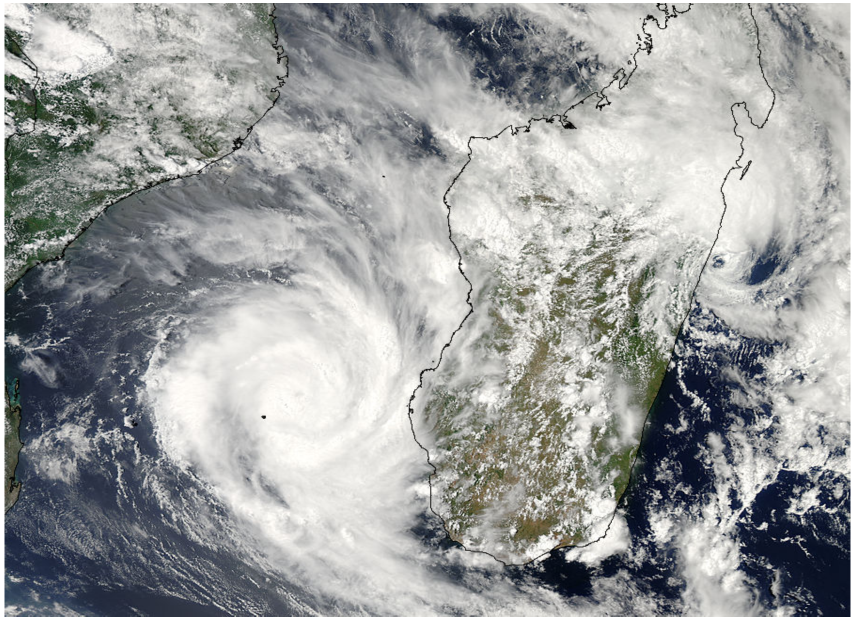
Hurricane Fanele, 20 January 2009

EDA based background error variance for Surface pressure



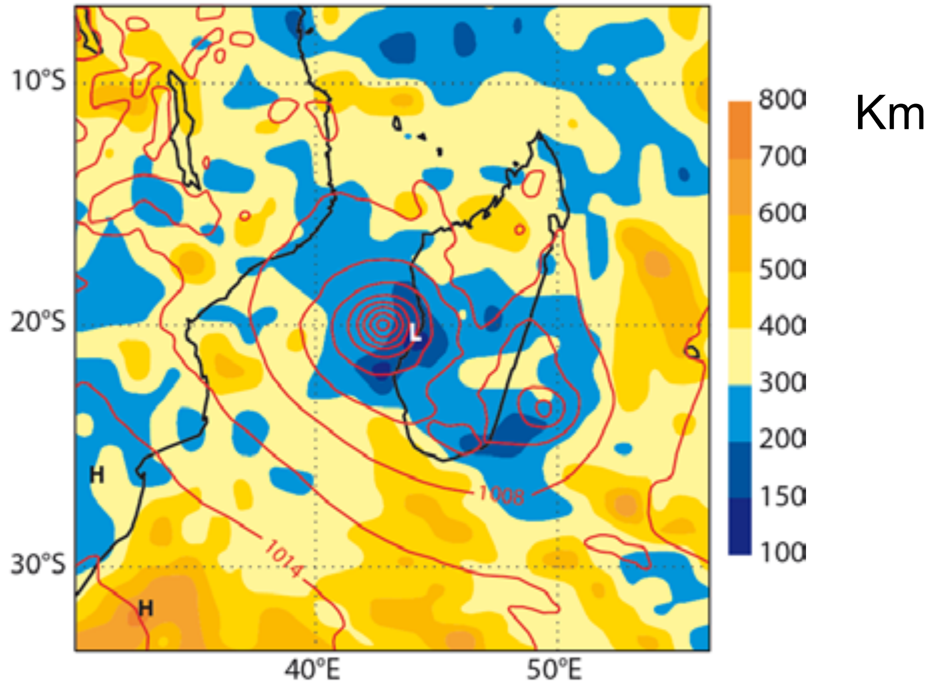
In November 2013 ECMWF will implement EDA based flow-dependent background error covariances in 4D-Var

The 25-member EDA has been used to estimate the background error covariance in 4D-Var.



Hurricane Fanele, 20 January 2009

EDA based background error covariance length scale for Surface pressure

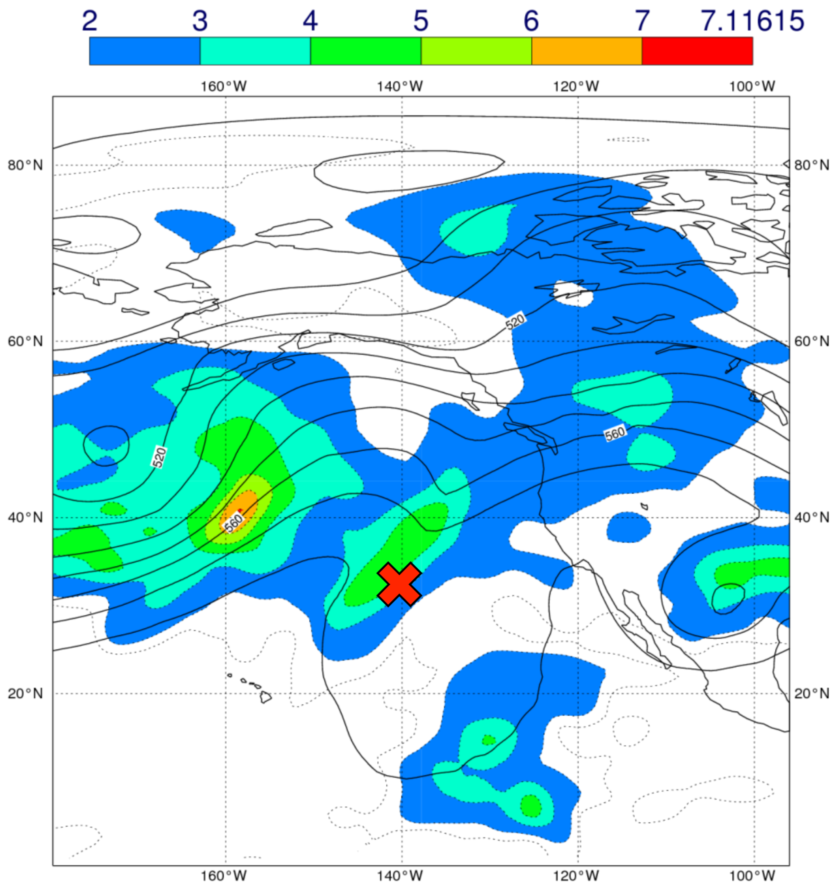


ECMWF

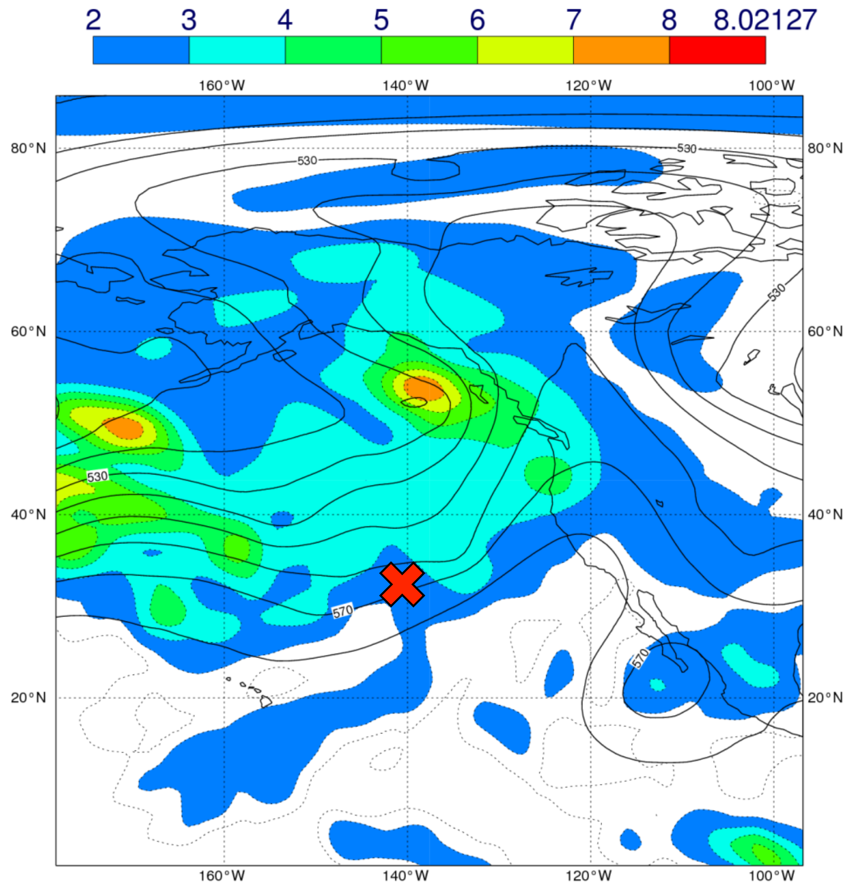
Why are flow-dependent covariances important?

Vertical correlations

2012-01-01 00Z

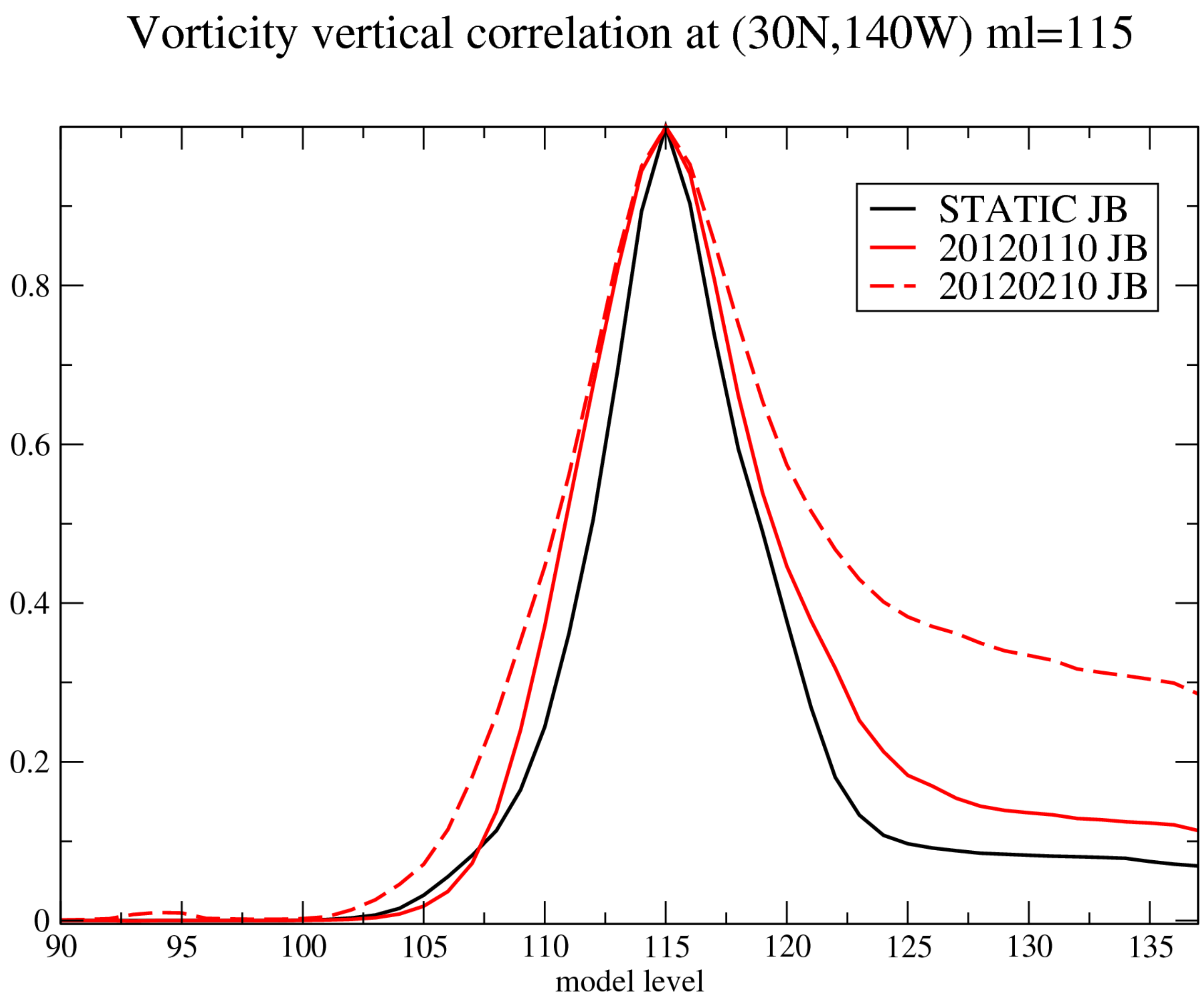


2012-02-01 00Z



Why are flow-dependent covariances important?

Vertical correlations



Why implementing Ensemble of Data Assimilations?

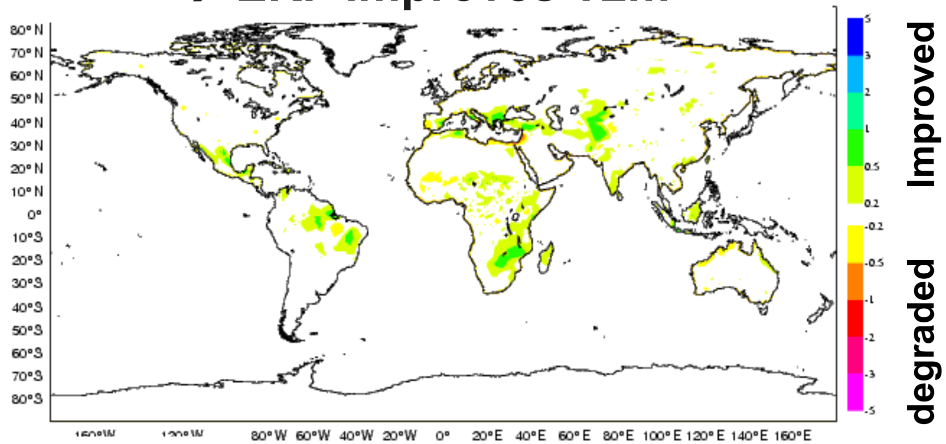
- **In general to estimate analysis and short range forecast uncertainty**
- To improve the initial perturbations in the Ensemble Prediction (implemented June 2010)
- To calculate static and seasonal background error statistics
- To estimate flow-dependent background error in 4D-Var - "errors-of-the-day" (implemented May 2011)
- To improve QC decisions and improve the use of observations in 4D-Var (implemented May 2011)
- To estimate flow-dependent background error covariances in 4D-Var - "errors-of-the-day" (implementation: November 2013)

Soil moisture assimilation using Extended Kalman Filter Implemented in November 2010

Impact on 2-metre Temperature

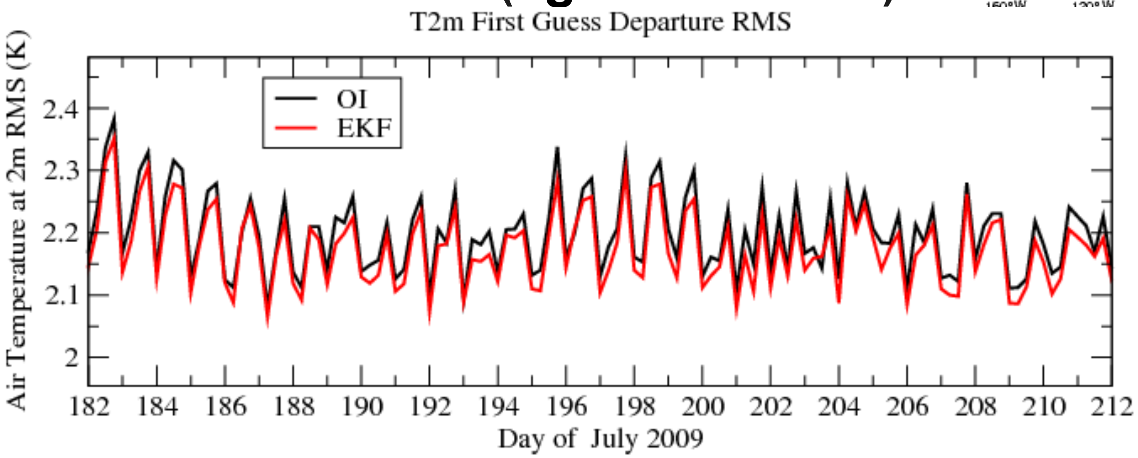
T2m error (OI-SEKF) 48h fc

→ EKF improves T2m



Compared to the old OI analysis, the simplified Extended Kalman Filter consistently improves T2m

Global mean RMS (against SYNOP)



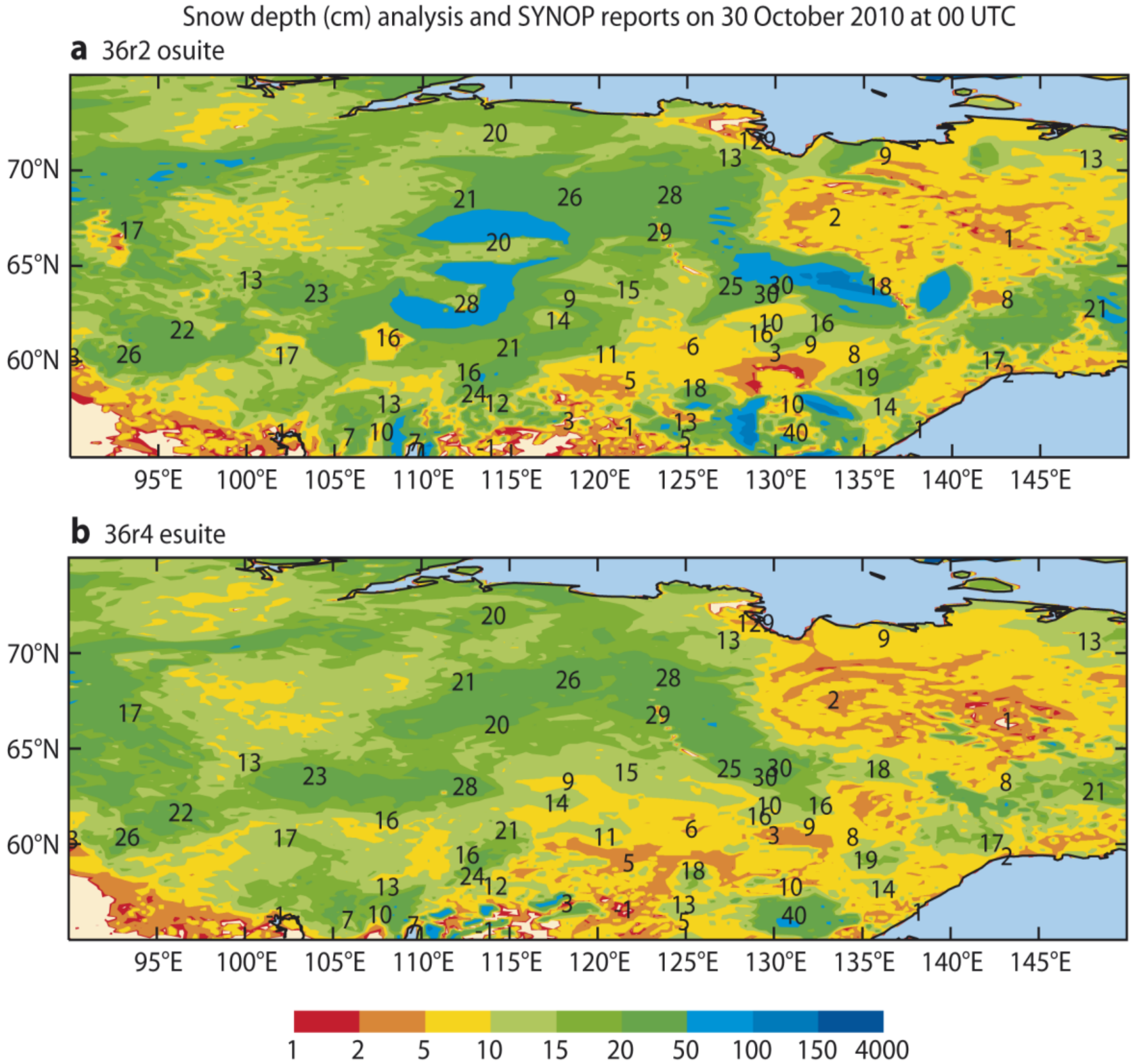
November 2010: new Optimum Interpolation snow analysis

The change improves the snow analysis significantly.

Spurious patterns and serious shortcomings of previous scheme resolved.

Better agreement with SYNOP snow observations.

- Top: Old Cressman using 24km NESDIS data
- Bottom: New OI (Brasnett 1999) approach using 4km NESDIS data



ECMWF

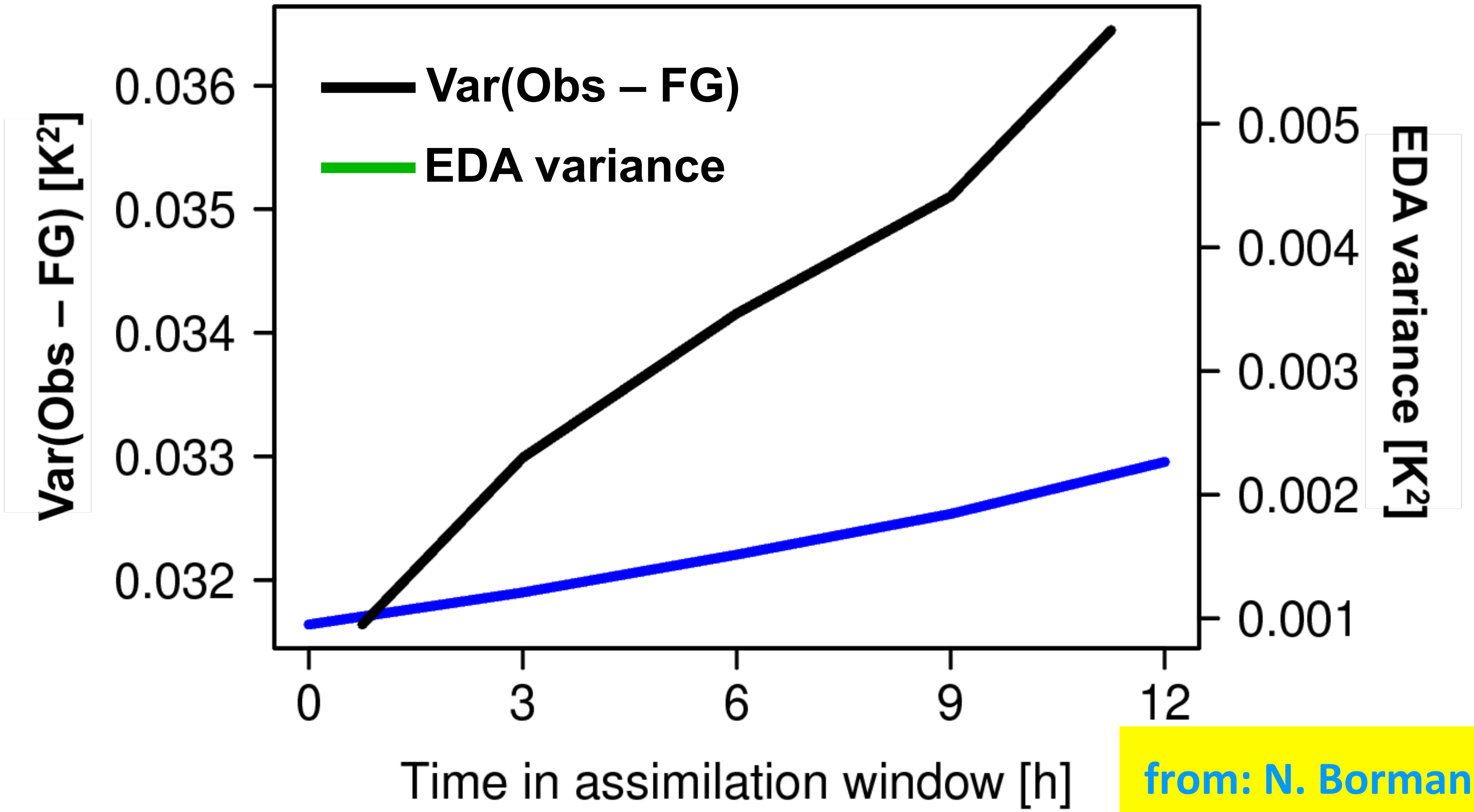
Coming soon in the data assimilation system at ECMWF

- Use of ASCAT data in EKF soil moisture analysis
- Introduction of cloud condensate in the data assimilation
- Retune observation errors for all data types
- A move to an Object Oriented Prediction System
- Improved scalability of 4D-Var

- Improvements to the EDA: model error parameterization, SST perturbations
- Improvements to the Jb formulation: anisotropic correlations, balance between q-T increments
- COPE – Continuous observation processing environment
- Long window 4D-Var: extend to 24 hour window

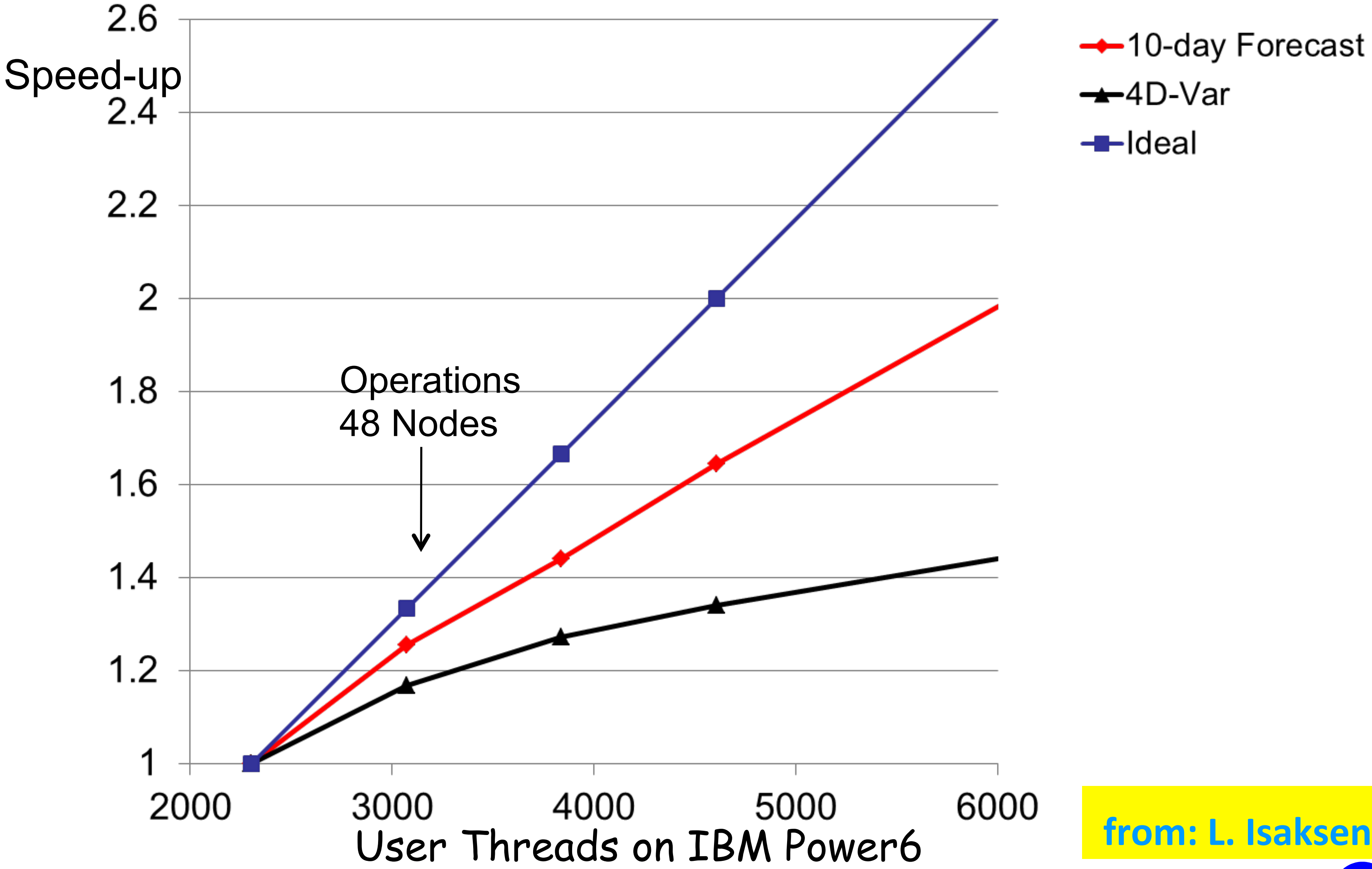
Evolution of EDA in assimilation-window

Example: AMSU-A, channel 8 (100-300 hPa)



from: N. Bormann

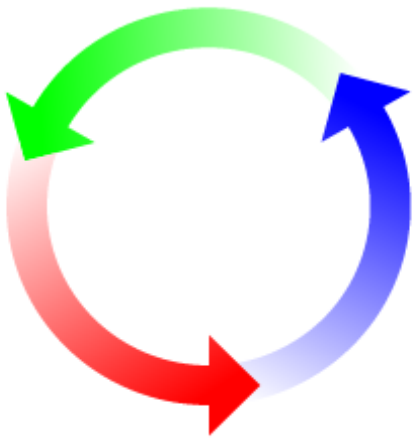
Scalability of T1279 Forecast and 4D-Var



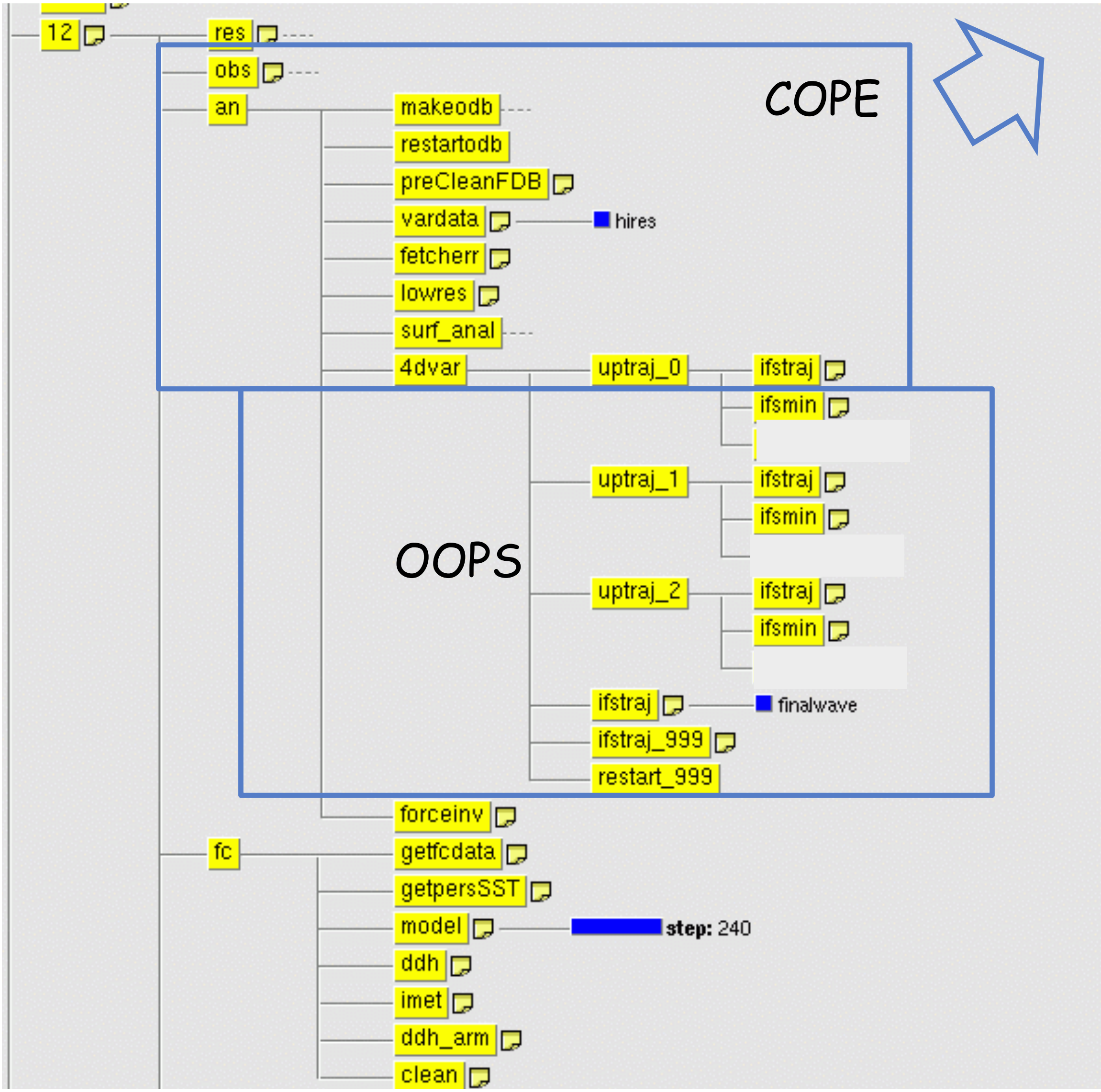
from: L. Isaksen

Continuous Observation Processing Environment (COPE)

- Implement a hub Observation Data Base (ODB) interface
- Shortens the time critical path by performing observation pre-processing and screening as data arrive
- Improve scalability by removing most observation related tasks from time critical path
- Reduce risk of failures in the analysis during the time critical path
- Enables near real-time quality control and monitoring of observations
- More modular software



Improving scalability of time critical Data Assimilation suite



COPE

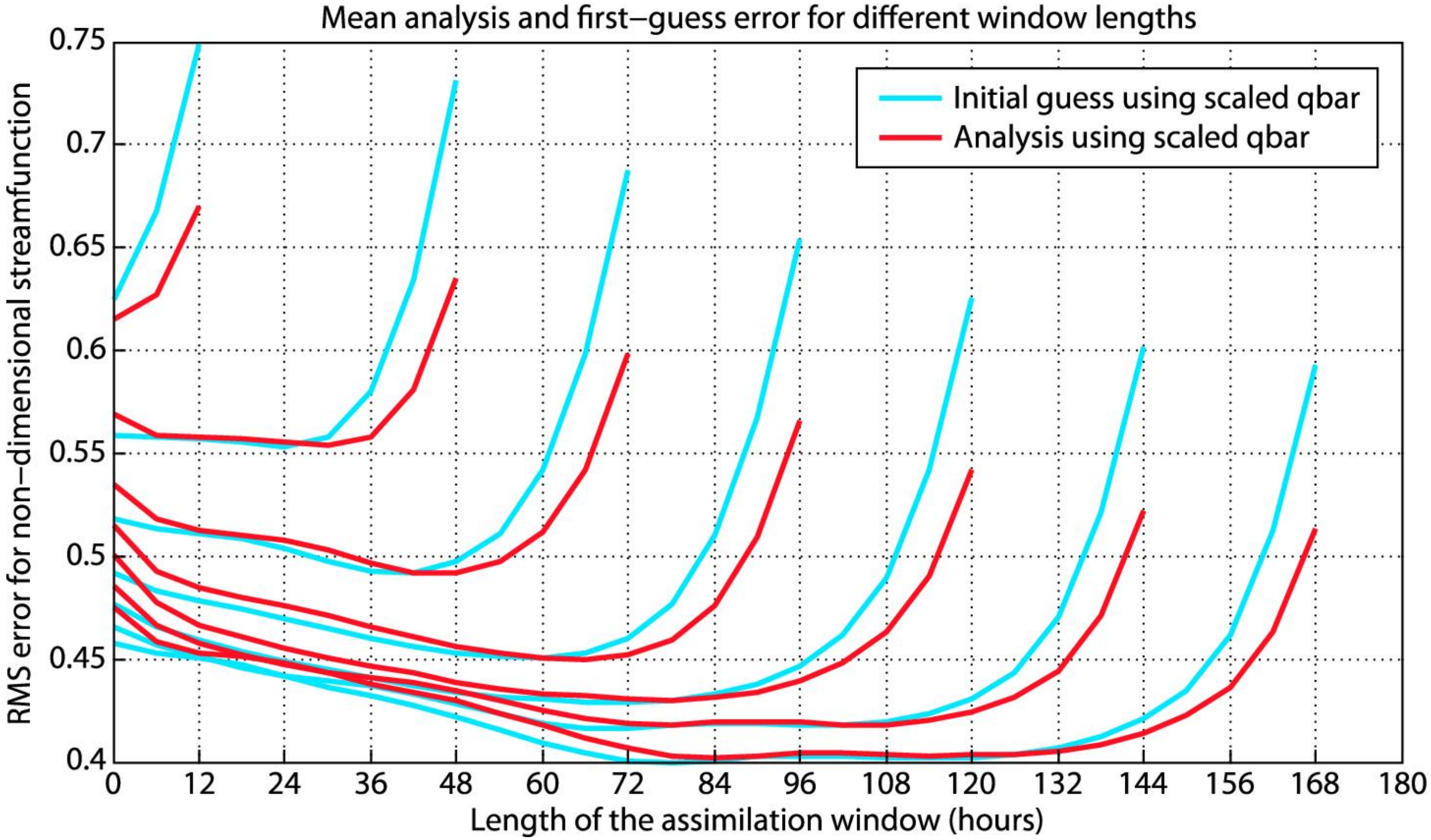
(Continuous Observation Processing Environment)

OOPS

(Object-Oriented Prediction System)

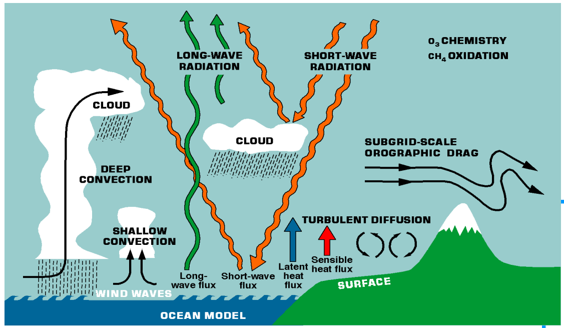
Long-window, weak-constraint 4D-Var - a longer-term project

Results based on a two-layer quasi-geostrophic model indicates that increasing the length of the analysis window is beneficial, even with a very simple model error representation.

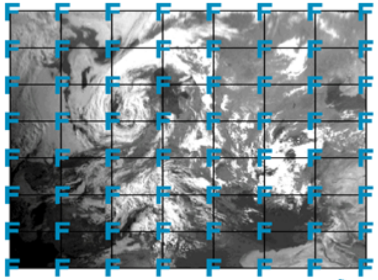


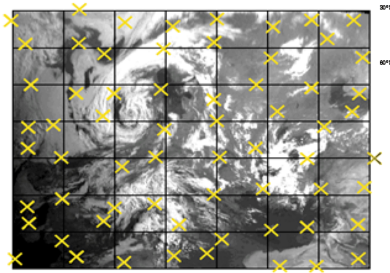
ECMWF

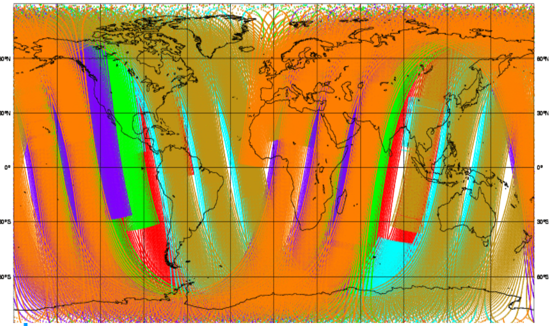
Data assimilation at ECMWF



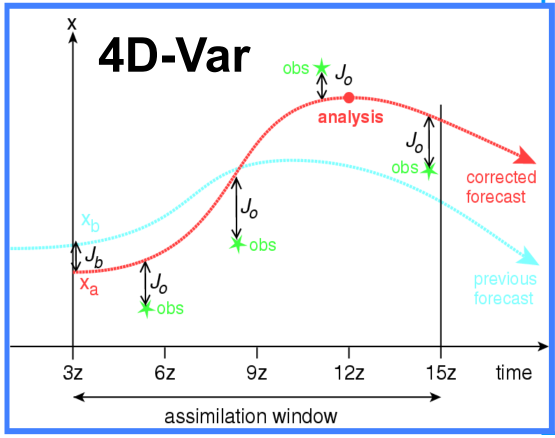
Forecast model

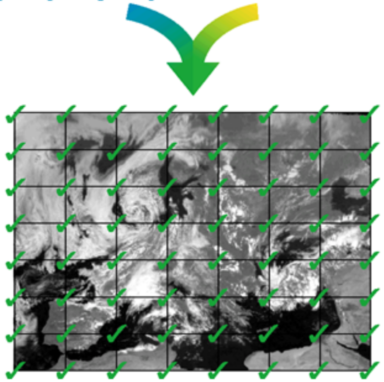


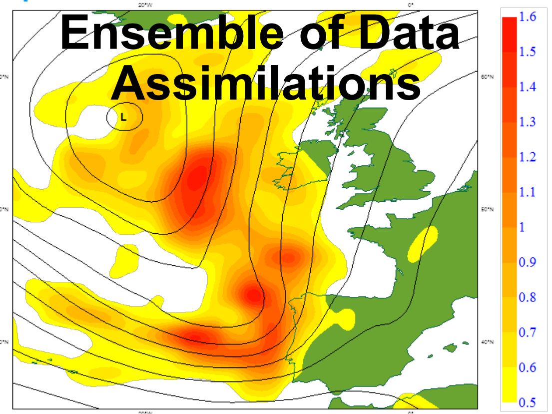


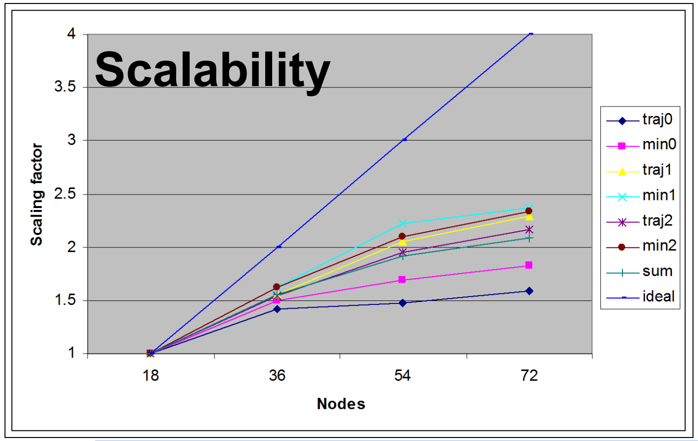


Observations









Methods

Progress and plans